SOME FAMILIES OF COUNT DISTRIBUTIONS FOR MODELLING ZERO-INFLATION AND DISPERSION

LOW YEH CHING

FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2016

SOME FAMILIES OF COUNT DISTRIBUTIONS FOR MODELLING ZERO-INFLATION AND DISPERSION

LOW YEH CHING

THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2016

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: Low Yeh Ching

Registration/Matric No: SHB090013

Name of Degree: Doctor of Philosophy

Title of Project Paper/Research Report/Dissertation/Thesis (“this Work”):

Some Families of Count Distributions for Modelling Zero-inflation and Dispersion

Field of Study:

Applied and Computational Statistics

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya (“UM”), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate’s Signature Date:

Subscribed and solemnly declared before,

Witness’s Signature Date:

Name:

Designation:

ABSTRACT

A popular distribution for the modelling of discrete count data is the Poisson distribution. However, count data usually exhibit over dispersion or under dispersion when modelled by a Poisson distribution in empirical modelling. The presence of excess zeros is also closely related to over dispersion. Two new mixed Poisson distributions, namely a three-parameter Poisson-exponentiated Weibull distribution and a four-parameter generalized Sichel distribution is introduced to model over dispersed, zero-inflated and long-tailed count data. Some of the theoretical properties of the distributions are derived and the distributions' characteristics are studied. A Monte Carlo simulation technique is examined and employed to overcome the computational issues arising from the intractability of the probability mass function of some mixed Poisson distributions. For parameter estimation, the simulated annealing global optimization routine and an EM-algorithm type approach for maximum likelihood estimation are studied. Examples are provided to compare the proposed distributions with several other existing mixed Poisson models. Another approach to modelling count data is by examining the relationship between the counts of number of events which has occurred up to a fixed time *t* and the inter-arrival times between the events in a renewal process. A family of count distributions, which is able to model under- and over dispersion, is presented by considering the inverse Gaussian distribution, the convolution of two gamma distributions and a finite mixture of exponential distributions as the distribution of the inter-arrival times. The probability function of the counts is often complicated thus a method using numerical Laplace transform inversion for computing the probabilities and the renewal function is proposed. Parameter estimation with maximum likelihood estimation is considered with applications of the count distributions to under dispersed and over dispersed count data from the literature.

ABSTRAK

Taburan Poisson merupakan suatu taburan yang popular untuk memodelkan data menghitung. Namun demikian, data menghitung biasanya memaparkan ciri di mana serakannya adalah melebihi atau kurang daripada apa yang dimodelkan oleh taburan Poisson. Kewujudan lebihan sifar juga adalah berkaitan dengan lebihan serakan ini. Dua taburan baru, iaitu taburan “Poisson-exponentiated Weibull” yang mempunyai tiga parameter dan taburan "generalized Sichel" yang mempunyai empat parameter dicadangkan untuk mengatasi masalah lebihan serakan, lebihan sifar dan ekor yang panjang. Sifat teoretikal dan ciri-ciri taburan baru ini dikaji. Suatu teknik simulasi Monte Carlo dikaji dan digunakan untuk mengatasi masalah perhitungan yang disebabkan oleh fungsi ketumpatan kebarangkalian taburan Poisson campuran yang hanya boleh ditulis dalam bentuk kamiran. Untuk tujuan anggaran parameter, rutin optimasi global "simulated annealing" dan pendekatan jenis "EM-algorithm" dikaji untuk anggaran "maximum likelihood". Contoh-contoh diberikan untuk membandingkan kesesuaian taburan baru ini untuk set data menghitung dengan taburan Poisson campuran yang lain. Pendekatan yang lain untuk memodelkan data menghitung adalah dengan mengkaji hubungan di antara bilangan kejadian yang telah berlaku sehingga suatu titik masa tetap t dengan jangka masa antara kejadian. Suatu famili taburan menghitung yang dapat memodelkan lebihan dan kurang serakan diperolehi apabila taburan "inverse Gaussian", konvolusi dua taburan gamma dan campuran taburan eksponensial digunakan sebagai taburan jangka masa antara kejadian. Fungsi kebarangkalian untuk data menghitung ini adalah rumit, maka satu kaedah yang menggunakan "numerical Laplace transform inversion" untuk menghitung kebarangkalian dan fungsi pembaharuan dicadangkan. Anggaran parameter melalui anggaran "maximum likelihood" dijalankan melalui aplikasi taburan-taburan tersebut dalam data menghitung lebihan dan kurang serakan.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor, Professor Dr. Ong Seng Huat for the support of my Ph.D study and for his expertise, patience, and understanding. His vast knowledge in many areas has immensely enriched my graduate experience and I appreciate his guidance in writing research work (journal and conference papers and this thesis). I would also like to thank my family for the support they provided me through my life. In particular, I must thank my husband for his support and encouragement, and my two children for their love and understanding throughout this graduate journey.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

AIC	:	Akaike Information Criterion
cdf	:	cumulative distribution function
EGIG	:	extended generalized inverse Gaussian
EM	:	Expectation-Maximization
GIG	:	generalized inverse Gaussian
ID	:	index of dispersion
iid	:	independent and identically distributed
ML	:	maximum likelihood
NB	:	negative binomial
pdf	:	probability density function
PIG	:	Poisson-inverse Gaussian
PGIG	:	Poisson-generalized inverse Gaussian
pdf	:	probability density function
pmf	:	probability mass function
$K_\nu(z)$	:	modified Bessel function of the third kind with index ν
Ω	:	parameter space
$\Re$	:	set of real numbers
$\mathbf{U}(.)$	:	score vector
$\mathbf{I}(.)$	:	expected information matrix
$\mathbf{J}(.)$	:	observed information matrix

LIST OF APPENDICES

CHAPTER 1: INTRODUCTION

1.1 Distributions for Statistical Modelling of Discrete Count Data

Discrete count data is encountered in many disciplines such as actuarial science, biology, computer science, engineering, linguistics, psychology, public health and sociology. Examples of applications of count data are the number of automobile insurance claims, citation counts and species abundance data. The binomial, Poisson and logarithmic distributions are some basic distributions for modelling discrete count data.

The well-known single parameter Poisson model is one of the basic models for discrete count data with infinite support. The Poisson distribution has probability mass function (pmf)

$$\Pr(X = k) = \frac{e^{-\theta}\theta^{k}}{k!}$$

for k = 0, 1, 2, ... and $\theta > 0$. The Poisson parameter θ also corresponds to the mean of the Poisson distribution. The Poisson distribution has a distinct characteristic in that its variance is equal to its mean. This characteristic is also known as equidispersion. A common issue when applying the Poisson distribution to model observed count frequency data is a violation of this variance-mean equality. To measure departure from equidispersion, the index of dispersion or Fisher dispersion index which is defined as $ID = \frac{\text{variance}}{\text{mean}}$ is commonly used. An equidispersed distribution such as the Poisson distribution will have an ID of value 1.

When the variance of the count data is larger than the mean, the index of dispersion is larger than 1 and the phenomenon is known as over dispersion. The presence of over dispersion can be attributed to, amongst others, unobserved heterogeneity in the data, clustering and small sample size. Cox (1983) showed that when certain requirements are fulfilled for the target parameter, maximum likelihood estimation in a simple model retains high efficiency under modest amounts of over dispersion. Lindsey (1999) argued that in regression modelling, over dispersion is present only when "the deviance is at least twice the number of degrees of freedom" (p. 560). Nevertheless, unaccounted over dispersion may cause problems such as biased inference and inefficient estimation in statistical modelling. It is also possible to have under dispersion, i.e. the variance is smaller than the mean though this is less common than over dispersion. In the case of under dispersion, the Fisher dispersion index takes value $0 < ID < 1$.

A popular approach to model under dispersion or over dispersion is by generalizing or extending the Poisson distribution. Mixed Poisson models, which are constructed by allowing the Poisson parameter to be a random variable with an appropriate probability structure, are intended for modelling latent heterogeneity in the population. Mullahy (1997) showed that unobserved heterogeneity, commonly assumed to be the source of over dispersion in count data modelling, have certain implications for the probability structures of such models. Examples of mixed Poisson distributions are the negative binomial (Greenwood & Yule, 1920), Poisson-inverse Gaussian (Holla, 1967; Sankaran, 1968) and the Delaporte distribution (Johnson, Kemp & Kotz, 2005). When over dispersion is present but there is no unobserved heterogeneity, one can consider other extensions to the Poisson distribution such as the generalized Poisson distribution defined by Consul and Jain (1973) and discussed in detail by Consul (1989), the double Poisson distribution proposed by Efron (1986), Poisson polynomial distribution (Cameron & Johansson, 1997), weighted versions of Poisson distributions (Castillo & Pérez-Casany, 2005) and the Conway-Maxwell-Poisson or COM-Poisson distribution (Conway & Maxwell, 1962; Shmueli, Minka, Kadane, Borle & Boatwright, 2005). Non-Poissonian approaches include Charlier series distribution (Ong, 1988) and its various generalizations (Ong, Chakraborty, Imoto & Shimizu, 2012), negative binomial mixture (Gómez-Déniz, Sarabia & Calderín-Ojeda, 2008), the Lagrangian Katz family of distributions (Gathy and Lefèvre, 2010) and non-parametric methods (Aitkin, 1996; 1999).

Over dispersion in observed count data is also closely related to presence of excess zeros. There are two types of zero counts that may occur in count data, i.e. structural zeros and sampling zeros. The difference between these two types of zeros can be clearly illustrated in an example from behavioural studies on alcohol abuse by He, Tang, Wang and Crits-Cristoph (2014). In a sample of observations on number of days that the subjects consumed alcohol in a study period, structural zeros are attributed to the existence of a subpopulation of subjects who does not drink alcohol at all, known as the non-risk group. Subjects who are at-risk (those who does consume alcohol) may still record a zero count response due to sampling and these zeros are known as sampling zeros. In some studies, it is necessary to distinguish between structural and sampling zeros in order to determine the different characteristics between the two groups. This is achieved by using a zero-inflated distribution (Johnson et al., 2005). Zero-inflated distributions split the zero counts into structural zeros and sampling zeros from a baseline distribution. If X is a random variable from the baseline distribution, its zero-inflated random variable Y is defined as

$$\Pr(Y = 0) = p + (1 - p)\Pr(X = 0)$$

$$\Pr(Y = k) = (1 - p)\Pr(X = k),\ k = 1, 2, 3, \ldots,$$

where $0 < p < 1$. For example, the zero-inflated Poisson distribution is defined as

$$P(X = 0) = p + (1 - p)e^{-\lambda}$$

$$P(X=k)=(1-p)\frac{e^{-\lambda}\lambda^{k}}{k!}, \quad k = 1, 2, 3, \ldots .$$

Another approach to address the structural and sampling zeros, with a subtle difference in interpretation, is by using a special case of the hurdle model first discussed by Mullahy (1986). A hurdle count data model with the hurdle at zero consists of a component which models the zero counts and another zero-truncated component distribution to model the nonzero observations. As such, the hurdle model interprets all the zero counts as structural zeros and the at-risk group is assumed to only produce nonzero positive counts. For example, Gurmu and Trivedi (1996) applied a hurdle model in modelling the number of recreational boating trips by a family in a year. If X is a random variable from the distribution of the nonzero counts, the probabilities of the random variable Y in a hurdle model are given as

$$\Pr(Y=0)=\gamma$$

$$\Pr(Y=k)=\frac{(1-\gamma)\Pr(X=k)}{1-\Pr(Y=0)}, \quad k = 1, 2, 3, \ldots .$$

In both of the approaches discussed, it is assumed that some (in the case of the zero-inflated distribution) or all (the hurdle model) of the excess zeros and the nonzero counts are not from the same data-generating process. If this assumption is not true, then the use of a zero-inflated distribution or hurdle model is not necessary. It will be of interest then to consider other alternatives that are able to model excess zeros as well as over dispersion.

In some disciplines such as computer science and linguistics, the observed count data may have a very long right tail. An example of long-tailed data is the number of citations for published journal articles (Zhu & Joe, 2009). A probability distribution is said to have a very long tail if the individual probabilities only become very small after a certain large k. The skewness of a distribution or the limiting ratio $\lim_{k\to\infty}\frac{\Pr(X=k+1)}{\Pr(X=k)}$ can be used to give an indication of the distribution's tail length. For the Poisson distribution, $\lim_{k\to\infty}\frac{\Pr(X=k+1)}{\Pr(X=k)}$ = 0, indicating that it has a short tail. Gupta and Ong (2005) have analysed the fit of some mixed Poisson distributions to model long-tailed count data.

Another approach to model discrete count data is by looking at the dual relationship between the occurrence of an event and the inter-arrival times between the events in a stochastic process. A counting process is a stochastic point process $\{N(t), t \geq 0\}$ where $N(t)$ represents the total number of events that have occurred by a fixed point in time t. The distribution of the event counts $N(t)$ is closely related to the distribution of the inter-arrival times between these events. Distributions that model the inter-arrival times are also known as duration models. A trivial example of this relationship is when the inter-arrival times are exponentially distributed. Then the counting process is a Poisson process with intensity $\lambda(t) = \lambda$ with pmf

$$\Pr\{N(t) = n\} = \frac{e^{-\lambda t}(\lambda t)^n}{n!}, \quad n = 0, 1, 2, \ldots .$$

The exponential distribution has a “memoryless” property due to its constant hazard function. Therefore, it is found to be inadequate in modelling duration data from many real applications. Continuous distributions with more flexible hazard functions such as the gamma and Weibull distributions are popular alternatives to the exponential distribution in duration analysis. Due to the interlinkage between event counts and the inter-arrival times (or duration) between events, it is then of interest to examine the count distributions arising from such non-exponential duration models. Research work in this area is further motivated by the intimate connection between the dispersion of the count distribution and the hazard function of the underlying duration models (Winkelmann, 1995). For example, McShane, Adrian, Bradlow and Fader (2008) have derived a count distribution with Weibull duration. In the same paper, McShane et al. (2008) also discussed other advantages of using a non-exponential duration model, which includes the ability to model heterogeneity.

1.2 Contributions of the Thesis

Part of the work in this thesis is motivated by the fact that there are a number of count frequency data sets with very high zero counts and/or very long right tails which may not be adequately fitted by existing mixed Poisson models. In the preceding section, we also contemplated on the necessity of a zero-inflated Poisson or hurdle model when there is presence of excess zeros in the data. An important result by Shaked (1980), which is aptly named as the Two-Crossings Theorem states that when a distribution is from the exponential family, its density and the density of an arbitrary mixture of this distribution with the same mean must 'cross' each other twice, from above in the first time, then from below in the second time. Based on this Two-Crossings Theorem, a mixed Poisson distribution, relative to the Poisson distribution, has a higher probability for the zero count and a longer right tail. This elevation of probability for zero counts and tail lengthening will vary according to the mixing distributions considered. As such, the choice of the mixing distribution is critical in order to obtain a distribution with high zero counts as well as a long right tail. We study and present two mixed Poisson distributions, namely the generalized Sichel distribution and the Poisson-exponentiated Weibull distribution. These two distributions are found to be more flexible and most importantly, fit better than other well-known mixed Poisson distributions when the count data has many zeros as well as a long tail. Since these distributions nest some well-known mixed Poisson distributions as special cases we also eliminate the need for a piecewise treatment in empirical count data modelling. A paper based upon this work has been submitted for publication.

We also address the computational hurdle encountered when evaluating the Poisson-exponentiated Weibull probabilities since it has an intractable probability mass function by using a Monte Carlo simulation technique. This computation issue has previously hindered the applications of many potentially useful mixed Poisson distributions, a well-known example being the Poisson-lognormal distribution. Statistical inference procedures for the generalized Sichel and Poisson-exponentiated Weibull distributions are also discussed. Karlis (2005) advocated an Expectation-Maximization (EM) algorithm for maximum likelihood estimation in mixed Poisson distributions. We adapt this EM-type algorithm for maximum likelihood estimation for the Poisson-Weibull distribution in particular. The procedures discussed can be extended and generalized for other mixed Poisson distributions. A paper based on this work is in progress.

Another motivation for the work in this thesis is the fact that in many real applications either the event count frequency data or the inter-arrival time between the event counts is recorded. The dual relationship between count distributions and their underlying duration models implies that given the knowledge of the count distribution, one can infer its underlying duration model, and vice versa. In econometrics, the inter-arrival times is a more familiar concept but such data may not be readily available. Moreover, the inter-arrival times need not be exponentially distributed. If such is the case and the observed count frequency data is available, one can exploit the interlinkage between the count distributions and the duration models to infer on the inter-arrival times. We derive and present a family of count distributions with inter-arrival times distributed as the inverse Gaussian, convolution of two gamma distributions and finite mixture of two exponential distributions. Due to the flexibility of the duration models' hazard functions, this family of distributions is able to model both over dispersion and under dispersion. Part of this work has resulted in two papers: One paper has been published (Ong, Biswas, Peiris & Low, 2015) and another paper will be published in a conference proceedings.

A major drawback in the applications of count distributions arising from non-exponential duration models is the computational issues on the count probabilities. These issues are attributed to numerical overflow caused by the infinite series or special mathematical functions in the probability mass function. In regard to this, we apply an efficient numerical inverse Laplace transform method based on the algorithm by Abate and Whitt (1992) to facilitate the computation of the count probabilities. The accuracy of the method is studied and found to be satisfactory. Applications of the count distributions arising from non-exponential duration models and the computation method are exemplified by fitting the models with real data from the literature. A paper based upon this work has been prepared for submission.

1.3 Organization of the Thesis

Chapter 2 contains a literature review on modelling of over dispersed, under dispersed and zero-inflated count data. A brief literature survey on the relevant statistical inference methods used to obtain the main findings given in Chapters 3 to 6 in this thesis is also provided.

The two new mixed Poisson distributions proposed in this thesis, namely the generalized Sichel distribution and Poisson-exponentiated Weibull distribution, are presented in Chapter 3. We study the shape of the distributions along with their characteristics in terms of skewness, length of tail and amount of zero-inflation.

Chapter 4 is concerned with the computation of probabilities and statistical inference for some mixed Poisson distributions. The expectation-maximization (EM) type algorithm for maximum likelihood estimation of mixed Poisson parameters is discussed. This chapter also contains a description of the hypothesis testing procedures for the two new mixed Poisson distributions discussed in Chapter 3. We show that the new mixed Poisson distributions give a superior fit to several real data sets selected from diverse fields in the literature.

A family of count distributions arising from non-exponential duration models are presented in Chapter 5. Apart from the probability mass function of the distributions, their characteristic with respect to modelling dispersion is discussed.

A major part of Chapter 6 is devoted to the numerical inverse Laplace transform method for computation of count probabilities arising from a renewal process with non-exponential duration. We propose an easily implemented and efficient method to compute the probabilities of the counts and subsequently the renewal function (expected number of renewals), given the Laplace transform of the inter-arrival times density function. The application of this method is illustrated on some existing and new count distributions in this context, along with model fitting on over dispersed and under dispersed data sets from the literature.

Finally, some concluding remarks are given in Chapter 7. An outline on future works is discussed at the end of this chapter.

CHAPTER 2: LITERATURE REVIEW

2.1 Statistical Distributions for Modelling Dispersion and Zero-inflated Count Data

The Poisson distribution is a benchmark model for modelling count data since it accounts for many inherent characteristics of count data such as positive skewness and zero counts. An exposition on the Poisson distribution and other univariate discrete distributions has been given by Johnson, Kemp and Kotz (2005). The Poisson assumption of equidispersion is often violated in observed count data, resulting in over dispersion or under dispersion. Solutions to overcome the presence of over dispersion or under dispersion include *ad hoc* methods, discretized continuous distributions, mixture models (for example, mixed Poisson models), generalizations of the birth and Poisson process, hurdle models, occurrence and duration dependent models. Kokonendji (2014) has presented a concise overview on count models for over dispersion and under dispersion. Some of these solutions are widely recognized in applied statistics. In their review on models for panel count data in insurance, Boucher and Guillén (2009) has discussed the use of mixed Poisson distributions, zero-inflated distributions and duration models.

The *ad hoc* methods are approaches such as the quasi-likelihood function (Wedderburn, 1974), extended quasi-likelihood (Nelder & Pregibon, 1987), combining quasi-likelihood estimation with maximum likelihood estimation (Brooks, 1984), pseudo-likelihood method (Carroll & Ruppert, 1988), simple likelihood method (Moore, 1986) and Efron's (1986) double exponential family. These methods do not assume a proper distribution for the count data.

In a mixed Poisson distribution, the Poisson parameter θ is allowed to be a random variable having an appropriate probability structure. Mixed Poisson distributions are always over dispersed relative to the simple Poisson distribution. The development of mixed Poisson distributions is closely linked to studies in accident-proneness and actuarial risk theory when accounting for different risk levels amongst individuals in an insurance portfolio. The earliest and simplest choice for the distribution of θ is the gamma density, resulting in a negative binomial distribution introduced by Greenwood and Yule (1920). Negative binomial regression models have been applied in diverse fields such as immunology (Periwal, Spagna, Shahabi, Quiroz & Shroff, 2005). Gupta and Ong (2004) proposed a generalized negative binomial distribution which has been found to fit some data sets better than the negative binomial distribution. The Delaporte distribution is obtained when the Poisson parameter follows a three-parameter gamma distribution (Johnson et al., 2005). Another popular mixed Poisson distribution is the Sichel distribution (Sichel, 1971) which is also known as the Poisson-generalized inverse Gaussian distribution. A special case of this distribution is the Poisson-inverse Gaussian (PIG) distribution (Holla, 1967; Sankaran, 1968). Hougaard, Lee and Whitmore (1997) considered the power variance mixture model, a large family of mixture distributions which includes the PIG as a special case. Kokonendji and Khoudar (2004) introduced the strict arcsine exponential dispersion model with pmf given as

$$\Pr(X=k)=\frac{A(k;1/\sigma^2)}{k!}\left(\frac{m^2\sigma^4}{1+m^2\sigma^4}\right)^{k/2}\exp\left[\frac{-1}{\sigma^2}\arcsin\sqrt{\frac{m^2\sigma^4}{1+m^2\sigma^4}}\right],\ k=0,\ 1,\ 2,\ \ldots,$$

where m is the mean and $\sigma^2=1/\alpha$, $\alpha>0$ being the parameter in the strict arcsine model introduced by Letac and Mora (1990). The function $A(.;.)$ is defined according to whether k is even or odd. Rigby, Stasinopoulos and Akantziliotou (2008) provided a general framework for the fitting of a family of mixed Poisson regression models by reparameterizing the mixing distributions to ensure that one of the parameters is always the mean of the mixed Poisson distribution. They also considered a Poisson-shifted generalized inverse Gaussian distribution but it is not found to fit the data better. Gómez-Déniz, Sarabia and Calderín-Ojeda (2011) developed a new unimodal two-parameter discrete distribution with a mode at zero count which is equally competitive with the negative binomial and PIG distribution in fitting over dispersed data in actuarial studies. The pmf of the distribution is given by

$$\Pr(X=k)=\frac{\log(1-\alpha\theta^{k})-\log(1-\alpha\theta^{k+1})}{\log(1-\alpha)},\ k=0,\ 1,\ 2,\ \ldots,$$

where $\alpha<1$, $\alpha\neq 0$ and $0<\theta<1$. Karlis and Xekalaki (2005) and Nikoloulopoulos and Karlis (2008) have given a review on some properties of mixed Poisson distributions. Xekalaki (2014) has discussed about over dispersion and studied the Waring distribution and its generalizations for modelling over dispersed count data. By studying the properties of the factorial cumulant generating function, Jørgensen and Kokonendji (2016) have proposed a class of discrete factorial dispersion models which includes some of the mixed Poisson distributions.

Gupta, Gupta, and Ong (2004) introduced the univariate and multivariate Poisson random effect models. In these models, the parameter θ of the Poisson distribution is modified by either adding (additive model) or multiplying (multiplicative model) with an unobserved random effect η. In turn, η can be modelled by a probability distribution with density function $g(\eta)$ such as the gamma distribution and the inverse Gaussian distribution. For the univariate additive model, the general pmf is defined as

$$\Pr(X=k)=\frac{e^{-\theta}\theta^{k}}{k!}\sum_{r=0}^{k}\int_{0}^{\infty}\frac{\binom{k}{r}}{\theta^{r}}e^{-\eta}\eta^{r}g(\eta)d\eta$$

whereas the general pmf for the univariate multiplicative model is given by

$$\Pr(X=k)=\frac{\theta^k}{k!}\int_0^\infty e^{-\theta\eta}\eta^k g(\eta)d\eta .$$

Cheng, Geedipally and Lord (2013) used a similar approach in developing the Poisson-Weibull generalized linear model for accident crash data.

Most generalizations of the Poisson distributions are able to accommodate both over dispersion and under dispersion. A well-known example is the generalized Poisson distribution proposed by Consul and Jain (1973) with pmf given as

$$\Pr(X=k)=\frac{e^{-(\theta_1+k\theta_2)}\theta_1(\theta_1+k\theta_2)^{k-1}}{k!}$$

such that $\Pr(X=k)=0$ for $k\geq m$ if $\theta_1+k\theta_2\leq 0$. It extends the Poisson distribution by including an additional parameter which can take positive, zero and negative values to account for over dispersion, equidispersion and under dispersion respectively. Castillo and Pérez-Casany (2005) considered a family of weighted versions of the Poisson distribution belonging to the exponential family. In general, a weighted Poisson distribution has pmf of the form

$$\Pr(X=k)=\frac{w(k)\Pr(X=k\,|\,\theta)}{E_\theta[w(X)]}$$

where $w(k)$ is nonnegative and $E_\theta[w(X)]$ is the mean with respect to the distribution of X depending on θ. Some $w(k)$ considered in the literature are $w(k)=\exp[rt(k)]$ (Castillo & Pérez-Casany, 2005), $w(k)=\exp[r\,|\,k-\mu\,|]$ (Ridout & Besbeas, 2004) and

$w(k)=\left(1+\sum_{t=1}^{p}\alpha_t k^t\right)^2$ (Cameron & Johansson, 1997).

Shmueli, Minka, Kadane, Borle and Boatwright (2005) have studied the statistical and probabilistic properties of the Conway-Maxwell-Poisson (COM-Poisson) distribution which is introduced by Conway and Maxwell (1962). The COM-Poisson pmf can be treated as a weighted Poisson distribution since its pmf is defined as

$$\Pr(X = k) = \frac{\theta^k}{(k!)^\nu} \frac{1}{Z(\theta,\nu)},$$

where $Z(\theta,\nu) = \sum_{j=0}^{\infty} \frac{\theta^j}{(j!)^\nu}$ for $\theta > 0$ and $\nu \geq 0$. Lord, Guikema, and Geedipally (2008) applied the COM-Poisson generalized linear model on accident data and found that the model performs equally well as compared to the negative binomial model. Sáez-Castillo and Conde-Sánchez (2013) proposed a regression model based on the hyper-Poisson distribution (Bardwell and Crow, 1964) as an alternative to the COM-Poisson and Poisson-Polynomial regression models.

Some researchers have taken a non-Poissonian approach in modelling under dispersion and/or over dispersion. For example, Jain and Consul (1971) proposed the generalized negative binomial distribution

$$\Pr(X = k) = \frac{n\Gamma(n+\beta k)}{k!\Gamma(n+\beta k - k + 1)} \alpha^k (1-\alpha)^{n+\beta k - k},$$

where $0 < \alpha < 1$, $|\alpha\beta| < 1$, $n > 0$ such that Pr(X = k) = 0 for $k \leq m$ if $n + \beta m < 0$. This generalized distribution nests the binomial and negative binomial distributions as special cases. Gómez-Déniz, Sarabia and Calderín-Ojeda (2008) proposed a negative binomial-inverse Gaussian distribution which is obtained by mixing one of the parameters in the negative binomial distribution with the inverse Gaussian distribution. Rodríguez-Avi, Conde-Sánchez, Sáez-Castillo, Olmo-Jiménez and Martínez-Rodríguez

(2009) developed a regression model based on the generalized Waring distribution (Irwin, 1968; Xekalaki, 1983), which is an extension of the negative binomial model and applied in accident theory. They compared its model fit with the negative binomial regression model.

2.1.1 Excess Zeros

Over dispersion is also closely related to the presence of excess zeros in the data. We say that there are excess zeros when the observed frequency of zero counts is significantly higher than the expected frequency predicted by an assumed model. Ridout, Demétrio and Hinde (1998) have given a review on the methods for modelling count data with excess zeros.

A natural model for the presence of excess zeros is the zero-inflated model. The pmf of a zero-inflated distribution has been given in Chapter 1. The zero-inflated model can be interpreted as a model for a mixture of two populations, and the zeros from the degenerate-at-zero distribution are known as structural zeros whereas those from the simple baseline model are sampling zeros. An example of a zero-inflated model is the zero inflated Poisson (ZIP) model, which can be considered as an extension of the simple Poisson distribution. The ZIP model has been used in regression modelling by Lambert (1992) in manufacturing, Böhning, Dietz, Schlattmann, Mendonca and Kirchner (1999) in dental epidemiology, Dalrymple, Hudson and Ford (2003) in a study on sudden infant death syndrome and Hall (2000) on horticulture data. Li et al. (1999) considered multivariate version of the ZIP models and found the models to be satisfactory in fitting real life data on manufacturing.

If count data exhibit both excess zero counts and over dispersion, the zero-inflated negative binomial (ZINB) distribution (Heilbron, 1994) will be more appropriate as a model. Jansakul and Hinde (2009) have derived a score test statistic for testing the NB against ZINB in regression modelling. Applications of the ZINB regression model can be found in the work by Yau, Wang and Lee (2003), Yip and Yau (2005), Trocóniz, Plan, Miller, & Karlsson (2009) and Ullah, Finch and Day (2010). Ridout, Hinde and Demétrio (2001) proposed a score test for testing the ZIP against ZINB alternatives and provided examples for cases with and without covariates. Phang and Ong (2006) proposed a zero-inflated inverse trinomial distribution as an alternative model to accommodate over dispersion and excessive zero counts in count data.

Through a series of simulation studies, Perumean-Chaney, Morgan, McDowall and Aban (2013) conceded that the zero-inflated distributions are necessary in modelling over dispersion and excess zeros. Of importance in this simulation study is that the data are generated from a zero-inflated distribution, thus the need exists to account for the two types of zeros is prevalent.

Gupta, Gupta and Tripathi (1996) introduced the zero adjusted generalized Poisson distribution where the possibility of zero-deflation, that is, the number of zeros is fewer than expected, is included.

2.1.2 Long-tailed distributions

Over dispersion is also related to the tail length of a discrete count data set. Classic examples of over dispersed and long-tailed data are the number of absenteeism among shift-workers (Arbous & Sichel, 1954), the distribution of Corbet’s Malayan butterfly with zeros (Bulmer, 1974) and fish species abundance data (Stein & Juritz, 1988). Ong and Muthaloo (1995) derived the modified Bessel function distribution of the third kind mixed Poisson (BF3-P) distribution for modelling very long-tailed data. Gupta and Ong (2005) pointed out that in a mixed Poisson distribution, careful consideration should be given to the choice of the mixing distribution in order to obtain a distribution with a longer tail than the negative binomial distribution. They analysed the fit of some mixed

Poisson distributions on long-tailed count data. Zhu and Joe (2009) derived a generalized Poisson inverse Gaussian family which is able to model long-tailed data, but computation of its probabilities require a recursion approach.

2.1.3 Count Distributions arising from Non-exponential Duration in a Renewal Process

In Chapter 1, we have briefly discussed the duality between event counts and inter-arrival times between the events in a stochastic process. Consequently, the modelling of count data can also be examined from the perspective of the duration between events, whereby the occurrence of an event leads to a count. When the sequence of inter-arrival times is independent and identically distributed, this is a special case known as a renewal process. A concise introduction to the theory of renewal processes and their basic properties can be found in the monograph by Cox (1962).

The hazard function is defined as $h(x)=\frac{f(x)}{1-F(x)}$ where $f(x)$ and $F(x)$ are the density function and cumulative distribution function of *X* respectively. A distribution is said to display negative duration dependence when $\frac{dh(x)}{dx}<0$ and positive duration dependence when $\frac{dh(x)}{dx}>0$. If the hazard function is monotonic, a direct relationship between the distribution's hazard function and its coefficient of variation can be established. Winkelmann (1995) has shown an important relationship between the behaviour of the duration model's hazard function and the dispersion of the count distribution in an underlying stochastic process. Duration models with increasing hazard function lead to under dispersed count distribution. On the other hand, duration models with a decreasing hazard function result in an over dispersed count distribution. Consequently, researchers have looked into count distributions arising from several non-exponential duration models. For example, Winkelmann (1995) has studied the Erlangian and gamma count distributions whilst McShane, Adrian, Bradlow and Fader (2008) have derived the Weibull count distribution for modelling under- and over dispersed count data. Lee (1996) has asserted the significance of this relationship and consequently developed a simulated likelihood approach based on the inter-arrival times for estimation of count data regression models. In the field of medical statistics, Lindsey (1998) has pointed out the importance of recording duration between events, such as bone fractures, as well as the frequency of the events due to presence of other factors such as switching treatments during the time period concerned. Zeviani, Ribeiro Jr., Bonat, Shimakura and Muniz (2014) applied the gamma count distribution in the context of regression modelling of under dispersed experimental data.

2.2 Statistical Inference

In this section, we provide a review on the statistical inference procedures used in the work for this thesis.

2.2.1 Maximum Likelihood Estimation

There are many parameter estimation methods for discrete distributions, for example, method of moments, M-estimation and minimum divergence estimation. One of the most popular methods is maximum likelihood (ML) estimation. Under regularity conditions, an ML estimator has many desirable properties such as efficiency, consistency and asymptotic normality. Moreover, an ML estimate is invariant under parameter transformation. Statistical properties of ML estimators are discussed by Casella and Berger (2002). ML estimation for discrete count distributions is performed as follows: Suppose X is a discrete count random variable which takes values k = 0, 1, 2, 3, …, with probability $\Pr(X = k)$. If the sample of interest consists of n independent and identically distributed (iid) observations $x_1, x_2, ..., x_n$ with unknown probability function, the likelihood function of the sample is

$$L(\boldsymbol{\omega} \mid x_1, x_2, ..., x_n) = \prod_{k=0}^{\infty} [\Pr(X = k)]^{f_k} ,$$

where f_k denotes the frequency of count k and $\boldsymbol{\omega} = (\omega_1, \omega_2, ..., \omega_m)^T$ the vector of unknown parameters for the assumed distribution and $\boldsymbol{\omega} \in \Omega \subset \Re^m$. Most of the time it is easier to work with the log-likelihood function instead, which is defined as

$$\log L(\boldsymbol{\omega} \mid x_1, x_2, ..., x_n) = \sum_{k=0}^{\infty} f_k \cdot \log \Pr(X = k) .$$

An ML estimator of $\boldsymbol{\omega}$ is the point at which $L(\boldsymbol{\omega} \mid x_1, x_2, ..., x_n)$ attains its maximum as a function of the parameters for the given sample. ML estimation determines the estimates of the unknown parameters by maximizing the sample's (log)-likelihood function. Therefore, ML estimates are defined as

$$\hat{\boldsymbol{\omega}} = (\hat{\omega}_1, \hat{\omega}_2, ..., \hat{\omega}_m)^T = \arg\max_{\boldsymbol{\omega}} \log L(\boldsymbol{\omega} \mid x_1, x_2, ..., x_n) .$$

The ML estimate is unique (provided it exists) if the parameter space Ω is convex and if the likelihood function is strictly concave in $\boldsymbol{\omega}$.

The first derivative of the log-likelihood function (also known as Fisher's score function) is defined as

$$\mathbf{U}(\boldsymbol{\omega}) = \frac{\partial \log L(\boldsymbol{\omega} \mid x_1, x_2, ..., x_n)}{\partial \boldsymbol{\omega}} .$$

From this definition, the score vector $\mathbf{U}(\boldsymbol{\omega})$ is simply the vector of first derivatives, taken with respect to the respective parameters of the assumed distribution.

The main task in ML estimation is to actually find the global maximum of the log-likelihood function. If the log-likelihood function is concave, the ML estimator can be found by setting Fisher's score function as being equal to zero. In a multiparameter setting, this may involve solving systems of nonlinear equations. In general, the log-likelihood may not possess such desirable properties and solving for Fisher's score function may not guarantee a global maximum. In practice, the log-likelihood function may even be too complicated or intractable to be solved analytically. To overcome the problem of finding the global maximum in such cases, direct maximization or a numerical optimization method is used to obtain the ML estimates. When using a numerical optimization method, one has to ensure that the algorithm converges to a global and not local maximum of the log-likelihood function. Two effective algorithms, namely the simulated annealing and the Expectation-Maximization algorithm, are used for the work in this thesis and are reviewed in the subsequent sections.

When regularity conditions are fulfilled, ML estimators are asymptotically normally distributed. Therefore, the variance and covariance of an ML estimate can be estimated by the corresponding elements in the inverse of the Fisher information matrix evaluated at the ML estimates. The Fisher information matrix is the matrix of second derivatives with its (i, j)-th element defined as

$$\{\mathbf{I}(\boldsymbol{\omega})\}_{ij} = E\left[\frac{\partial \log L}{\partial \omega_i}\frac{\partial \log L}{\omega_j}\right] = -E\left[\frac{\partial^2 \log L}{\partial \omega_i \omega_j}\right].$$

Consequently, the standard errors of the ML estimates are taken to be the square roots of the diagonal elements of $\mathbf{I}(\hat{\boldsymbol{\omega}})^{-1}$. In the event that the expected information matrix is intractable, one can instead use the observed information matrix

$$\{\mathbf{J}(\boldsymbol{\omega})\}_{ij} = \left[-\frac{\partial^2 \log L}{\partial \omega_i \omega_j} \right].$$

The observed information matrix is simply the negative of the matrix of second derivatives (the Hessian matrix) of the log-likelihood. Efron and Hinkley (1978) advocated the use of the observed information matrix in place of the expected information matrix. When the derivatives cannot be obtained analytically, finite difference methods can be used to approximate the derivatives in the Hessian matrix.

If the sample size is small, standard errors of the ML estimates may be estimated using bootstrap methods (Efron, 1981).

2.2.2 Simulated Annealing

Simulated annealing (Kirkpatrick, Gelatt & Vecchi, 1983; Corana, Marchesi, Martini & Ridella, 1987) is a very robust algorithm for finding the global maximum of a function. The use of numerical optimization in the maximum likelihood problem must proceed with care to avoid undesirable outcomes such as slow or even non-convergence and inability to cope with difficult functions with ridges and plateaus. Although the use of different starting values may resolve some of these problems, there are still uncertainties at stake. On the other hand, the simulated annealing algorithm has been found to work well even in high-dimensional, non-quadratic and non-smooth log-likelihood functions with many local maxima (Goffe, Ferrier & Rogers, 1994).

The concept of the simulated annealing algorithm originates from the cooling of molten metal in thermodynamics. Annealing means ‘slow cooling’. In the cooling process of molten metal, random fluctuations in energy allows the metal’s energy state to escape local minima to achieve the global minimum. Simulated annealing works by drawing parallels between minimizing a function and the metal’s annealing system. In the maximum likelihood problem, the algorithm works by exploring the entire surface of the negative of the sample's log-likelihood function and searches for the optimum value while moving both uphill and downhill. Consequently, it is independent of starting values and is able to move out of local maxima to achieve the global maximum.

Critical starting parameters of the simulated annealing algorithm are the initial temperature, the starting vector of parameters and the step length for the vector of parameters. The algorithm starts by moving with large step lengths to get an overview of the log-likelihood function's surface. As the temperature and step length decreases, it will then gradually focuses on the most possible area for the global maximum and at the same time taking downhill moves to escape local maxima. The only drawback to this algorithm is in its longer execution time. Corana et al. (1987) has recommended some input values for the algorithm's parameters. Building upon this recommendation, a strategy to optimize the algorithm's performance by selecting appropriate parameter inputs is given by Goffe et al. (1994).

2.2.3 Expectation-Maximization (EM) Algorithm

The Expectation-Maximization (EM) algorithm is an iterative algorithm first introduced by Dempster, Laird and Rubin (1977) for ML estimation when the observations are seen as incomplete data. The iterative algorithm derives its name from the two steps involved in each of the iterations, namely an expectation step followed by a maximization step. In using this approach, one formulates the problem by first visualizing that there exists two sample spaces Y and X and a many-to-one mapping from X to Y. The observed data points $\mathbf{y}$ are a realization from Y. The corresponding $\mathbf{x}$ from X, referred to as the complete data (though it can actually be the parameters), cannot be observed directly and must be inferred from $\mathbf{y}$. The relationship between the complete-data specification and the incomplete-data specification is given by

$$g(\mathbf{y}\,|\,\boldsymbol{\varphi}) = \int_{X(\mathbf{y})} f(\mathbf{x}\,|\,\boldsymbol{\varphi})d\mathbf{x}$$

where $f(\mathbf{x}\,|\,\boldsymbol{\varphi})$ is a family of sampling densities depending on parameters $\boldsymbol{\varphi}$ and $g(\mathbf{y}\,|\,\boldsymbol{\varphi})$ are its corresponding family of sampling densities. Given the observed data points **y,** EM algorithm maximizes $g(\mathbf{y}\,|\,\boldsymbol{\varphi})$ with respect to $\boldsymbol{\varphi}$, through the associated family $f(\mathbf{x}\,|\,\boldsymbol{\varphi})$. In general, the EM algorithm proceeds as follows (Dempster et al., 1977).

Define the function $Q(\boldsymbol{\varphi}'\,|\,\boldsymbol{\varphi}) = E[\log f(\mathbf{x}\,|\,\boldsymbol{\varphi}')\,|\,\mathbf{y},\boldsymbol{\varphi}]$. At the *k*-th iteration of the algorithm:

E-step: Compute the $Q(\boldsymbol{\varphi}\,|\,\boldsymbol{\varphi}^{(k)}) = E[\log f(\mathbf{x}\,|\,\boldsymbol{\varphi})\,|\,\mathbf{y},\boldsymbol{\varphi}^{(k)}]$.

M-step: Compute $\boldsymbol{\varphi}^{(k+1)}$ by maximizing the function $Q(\boldsymbol{\varphi}\,|\,\boldsymbol{\varphi}^{(k)})$.

The iterations are terminated by a pre-determined stopping criterion such as one based on the relative change of the log-likelihood functions. Stochastic versions of the EM algorithm such as stochastic EM (Celeux & Diebolt, 1985) and Monte Carlo EM (Wei & Tanner, 1990) are introduced when the computation in the E-step is intractable. Celeux, Chauveau and Diebolt (1995) have examined the characteristics and relationships of these variants of the EM algorithm. On the other hand, modifications to an intractable M-step can be made through the introduction of a numerical optimization method, resulting in the ECM algorithm (Meng & Rubin, 1993), amongst others.

The attractiveness of the EM algorithm and its variants for ML estimation has resulted in it being adapted and applied in various contexts. For example, Chan and Ledolter (1995) on time series models involving counts, McLachlan (1997) on modifications to generalized linear models for handling over dispersed count data, Balakrishnan and Pal (2012) for cure rate models and so on. Other than finding ML estimates, the EM algorithm also provides useful by-products such as posterior expectations for predicting future outcomes in mixed Poisson regression models (Karlis, 2001) and observed information matrix (Louis, 1982).

There are some limitations when using the EM algorithm in ML estimation. Karlis and Xekalaki (2003) highlighted some of these issues, such as the algorithm's high dependency on starting values, suitability of the stopping criterion, slow convergence and convergence to local instead of global optimum. There are many research work dedicated to improving the EM algorithm, for example Louis (1982), Jank (2005), Lange (1985) and so on.

2.2.4 Goodness-of-fit

The most common procedure for testing distributional assumptions in the discrete case is the chi-square goodness-of-fit test introduced by Pearson (1900). The hypotheses for this test could be formulated as:

H_0 : The data follow a specified distribution with *m* parameters.

H_A : The data do not follow a specified distribution with *m* parameters.

This chi-square goodness-of-fit test is independent of the form of the distribution being tested. Using this goodness-of-fit test on discrete count data, observed values are divided into *t* mutually exclusive and exhaustive classes. For each class $i = 1, 2, \ldots, t$, the observed frequency O_i is compared to the expected frequency E_i under H_0. The resulting test statistic is defined as

$$\chi^2 = \sum_{i=1}^{t} \frac{(O_i - E_i)^2}{E_i} .$$

The test statistic has an approximate chi-square distribution with $t - 1 - m$ degrees of freedom. A conservative rule for this approximate distribution to be applicable is to ensure that each E_i is at least 5. Based on a power study of the test, Cochran (1954) relaxed this condition by recommending the E_i's to be at least 1. The test statistic quantifies the difference between the observed frequency and the frequency that should be observed under the specified distribution. Therefore, intuitively the larger the test statistic, the poorer is the fit of the specified distribution.

2.2.5 Model Selection

In empirical modelling, the Akaike information criterion (AIC) (Akaike, 1974) and Bayes' information criterion (BIC) (Schwarz, 1978) are commonly used to determine the best amongst competing candidate models in approximating the unknown true model. Based on the idea of the Kullback-Leibler distance between the "true model" and a candidate model and its maximized log-likelihood function, the AIC is defined as

$$\mathrm{AIC} = -2\log(L(\hat{\boldsymbol{\omega}} \mid x_1, x_2, \ldots, x_n)) + 2m,$$

where m is the number of estimable parameters in the maximum likelihood estimation for the model. The AIC penalizes models with more parameters, in line with the concept of parsimony.

A small-sample variant of AIC, known as AIC_c is studied by Hurvich and Tsai (1989) and can be written as

$$\mathrm{AIC}_c = \mathrm{AIC} + \frac{2m(m+1)}{n-m-1},$$

where n is sample size. In determining which criterion to be used, Burnham and Anderson (2004) recommended the use of AIC_c when the ratio n/m is less than 40, using the value of m from the highest-dimensioned model. The use of either AIC or AIC$_c$ must be consistent in the analysis. In choosing between the competing models, only a difference in AIC or AIC$_c$ values of more than 5 is to be considered as substantial (Burnham & Anderson, 2004).

On the other hand, for a model with sample size n the BIC is calculated as

$$\text{BIC} = -2\log(L(\hat{\boldsymbol{\omega}} \mid x_1, x_2, ..., x_n)) + m\log(n) \text{ .}$$

The BIC is more sensitive towards the model's dimension especially when the sample size is large. The model with the smallest AIC or BIC value is selected as the best model as it is interpreted as one that is "closest" to the unknown true model which generated the data.

In the event that all the competing models are far from the unknown true model, an alternative is to use Takeuchi's information criterion (TIC) (Takeuchi, 1976) if the sample size is quite large. The TIC is computed as

$$\text{TIC} = -2\log(L(\hat{\boldsymbol{\omega}} \mid x_1, x_2, ..., x_n)) + 2 \cdot \text{tr}(\mathbf{U}(\hat{\boldsymbol{\omega}})\mathbf{I}(\hat{\boldsymbol{\omega}})^{-1}) \text{ ,}$$

where $\mathbf{U}(\boldsymbol{\omega})$ and $\mathbf{I}(\boldsymbol{\omega})$ are the first and second mixed partial derivatives of the log-likelihood function, respectively. Reliable estimation of the elements in these matrices is crucial but difficult especially when the model is of high dimension, unless the sample size is very large. Due to the possible estimation error of the two matrices, unless the sample size is very large, the use of TIC is not advocated (Burnham & Anderson, 2002).

2.2.6 Hypothesis Testing

Hypothesis testing involves the problem of deciding which of the two hypotheses denoted by H_0 the null hypothesis and H_A the alternative hypothesis is true. The hypotheses are statements regarding the value or values taken by a population parameter or parameter vector. For the work in this thesis, the models in H_0 and H_A are nested. The model represented by H_0 is the more restrictive model that can be derived as a special case of the model represented by H_A. In this case, three well-known procedures for hypothesis testing are the likelihood ratio test (Neyman & Pearson, 1928), Rao's score test (Rao, 1948) and Wald test (Wald, 1943). All three tests are asymptotically equivalent but only the likelihood ratio test and score test are invariant under transformation of the parameters. Buse (1982) gives a concise account on these three tests and some of their properties. For the work in this thesis, we use the likelihood ratio test and the score test for hypothesis testing.

The likelihood ratio test statistic is constructed by using a ratio of likelihood functions. The log-likelihood function equivalent of the test statistic is defined

$$T_{LR} = -2[\log(L(\hat{\boldsymbol{\omega}} \mid x_1, x_2, ..., x_n)) - \log(L(\hat{\boldsymbol{\omega}}_0 \mid x_1, x_2, ..., x_n))],$$

where $\hat{\boldsymbol{\omega}}$ and $\hat{\boldsymbol{\omega}}_0$ are the vector of ML estimates evaluated at the alternative hypothesis and null hypothesis, respectively. In this special case, T_{LR} is asymptotically chi-square distributed with 1 degree of freedom. When the parameter of interest lies on the boundary of the parameter space under the null hypothesis, the distribution of T_{LR} is a mixture of the chi-square distribution and a distribution degenerate at zero (Self & Liang, 1987).

Rao's score test is also known as the Lagrange Multiplier test in econometrics. An extensive explanation on the score test can be found in the monograph by Cox and Hinkley (1974). Rao (2005) has given a historical review and some recent developments on the score test. The score test has the advantage of requiring only the estimates of the parameters under the null hypothesis, which are often simpler or require less time to compute. The score test statistic is defined as

$$T_{RS} = \mathbf{U}^T(\hat{\boldsymbol{\omega}}_0)\mathbf{I}^{-1}(\hat{\boldsymbol{\omega}}_0)\mathbf{U}(\hat{\boldsymbol{\omega}}_0),$$

where $\hat{\boldsymbol{\omega}}_0$ is the vector of maximum likelihood estimates under the null hypothesis. The score test statistic T_{RS} is also asymptotically chi-square distributed with 1 degree of freedom. One could also use the observed information matrix $\mathbf{J}(\boldsymbol{\omega}) = \left[-\frac{\partial^2 \log L}{\partial \omega_i \omega_j} \right]$ in place of the Fisher information matrix but Morgan, Palmer and Ridout (2007) cautioned on obtaining a negative score test statistic. Freedman (2007) suggested computing the information matrix using the unrestricted ML estimates to ensure the consistency of the score test statistic. Authors in the literature have used both forms. Broek (1995) used the expected Fisher information matrix to compute the score test statistic for testing zero-inflation while Atkinson and Yeh (1982) used the observed information matrix in their test for the significance of the Sichel distribution's additional parameter. Gupta, Gupta and Tripathi (2005) used the expected Fisher information except for one of the elements in the matrix, whereby the computation is not feasible hence the expectation operator was omitted.

CHAPTER 3: SOME MIXED POISSON DISTRIBUTIONS

3.1 Introduction

Observed data in empirical modelling is often over dispersed, violating the Poisson distribution's assumption of equidispersion. One of the solutions to model over dispersion is by use of mixed Poisson distributions. A mixed Poisson distribution is constructed by taking the Poisson parameter θ to be a random variable with an appropriate probability structure. In this context, the distribution of θ is known as the mixing distribution. This approach accounts for unobserved heterogeneity in the population, which has been identified as one of the causes of over dispersion.

The Poisson-generalized inverse Gaussian (PGIG) or Sichel distribution (Sichel, 1971) is a popular mixed Poisson distribution obtained when θ is a generalized inverse Gaussian (GIG) random variable. The Sichel distribution is a long-tailed distribution that is found to be suitable for highly skewed data and it has been used, amongst others, to model insurance claim counts, protein abundance, word frequency in a text and consumer purchase behaviour. Based on Shaked's (Shaked, 1980) Two-Crossings Theorem, a mixed Poisson distribution, relative to the Poisson distribution, has a higher probability for the zero count and a longer right tail. Willmot (1990) highlighted similarities between the right tail of the mixing distribution and that of the resulting mixed Poisson distribution. With these points in mind, in this chapter we derive a new mixed Poisson distribution by considering the extended generalized inverse Gaussian (EGIG) distribution as the mixing distribution. We name this new distribution as the generalized Sichel distribution. The EGIG distribution is derived by Jørgensen (1982) as a power transformation of the GIG random variable. As such, the EGIG distribution is more versatile with one additional parameter than the GIG distribution. Gupta and Viles (2011) has shown that this flexible distribution for analysing lifetime data is able to fit some data sets better than the GIG distribution. Statistical inferences of the EGIG distribution such as parameter estimation with maximum likelihood estimation and hypothesis testing have been studied by Gupta and Viles (2011). The generalized Sichel distribution inherits the versatility of its mixing distribution and we show that this new mixed Poisson distribution is able to model over dispersed, zero-inflated and long-tailed count data sets through applications with simulated and real data sets. This generalized Sichel distribution nests the negative binomial (NB), Poisson-inverse Gaussian (PIG) and Sichel distributions as special cases. This is a useful feature in empirical modelling because the generalized Sichel distribution eliminates the need of piecewise treatment when fitting a data set that is believed to have the NB, PIG or Sichel distribution.

Recently, Cheng, Geedipally and Lord (2013) proposed the Poisson-Weibull generalized linear model in the modelling of automobile crash count data. Although a regression model is appealing in that it is able to incorporate explanatory variables, a simple empirical model can be useful to describe the observed count data when such explanatory variables are unclear or yet to be determined. A similar point of view is given by Ridout and Besbeas (2004). In this chapter, we also examine the mixed Poisson-exponentiated Weibull distribution as an extension to the Poisson-Weibull distribution for modelling over dispersed discrete count data without covariates. There is a wealth of literature on extensions to the Weibull distribution. One of them is the exponentiated Weibull distribution proposed by Mudholkar, Srivastava and Freimer (1995). The exponentiated Weibull distribution is a generalization of the Weibull family and nests the exponential, exponentiated exponential and Weibull distributions.

The remaining of this chapter is organized as follows. In the next section, we give a survey on the existing literature on mixed Poisson distributions. The generalized Sichel distribution and Poisson-exponentiated Weibull distribution are presented in subsequent sections. In Section 3.5, the flexibility of the shape of the proposed distributions is illustrated through their probability plots. We also examine the characteristics of the distributions with respect to the dispersion, zero-inflation and the third central moment inflation indices in Section 3.6. Finally, concluding remarks are given in Section 3.7.

3.2 Literature Review

In the existing literature on mixed Poisson distributions, various authors have proposed different candidate distributions as the mixing distribution. One of the earliest papers is by Greenwood and Yule (1920) who used the gamma distribution as the mixing distribution, resulting in the well-known negative binomial (NB) distribution. The pmf of the NB distribution is

$$\Pr(X=k)=\binom{\alpha+k-1}{\alpha-1}\left(\frac{\beta}{\beta+1}\right)^{k}\left(\frac{1}{\beta+1}\right)^{\alpha} \tag{3.2.1}$$

for k = 0, 1, 2, ..., and α, β > 0. With only two parameters and a closed form expression for its pmf, the NB distribution is relatively simple and widely used in modelling over dispersed count data. Both the Poisson and NB distributions belong to the natural exponential family with quadratic variance functions, which enjoy many useful properties (Morris, 1982). Lawless (1987) did a study on the statistical inference of the NB regression model using the multiplicative random effects approach.

When the mixing distribution is the inverse Gaussian distribution, the resulting Poisson mixture is known as the Poisson-inverse Gaussian (PIG) distribution. This distribution is proposed by Holla (1967) and Sankaran (1968), independently. The pmf of the PIG distribution is given as

$$\Pr(X=k)=\sqrt{\frac{2\alpha}{\pi}}\,\frac{\exp(\alpha\sqrt{1-\theta})\left(\frac{\alpha\theta}{2}\right)^{k}}{k!}\,K_{k+\frac{1}{2}}(\alpha) \tag{3.2.2}$$

for k = 0, 1, 2, ..., and $\alpha > 0$, $0 < \theta < 1$ and $K_\nu(z)$ is the modified Bessel function of the third kind with index ν. This distribution has been found to be useful in modelling long-tailed data such as species abundance data (Ord & Whitmore, 1986) and insurance claim data (Tremblay, 1992). Shaban (1981) provided a recurrence relation for PIG probabilities to increase the computation speed and efficiency. However, this approximation is in general very poor for large k. In view of this, Ong (1998) has re-written the PIG distribution in terms of a mixed Poisson distribution with the inverted gamma as mixing distribution and showed that the approximation improves with large counts, hence providing a better approximation for the tail probabilities. Ong (1998) also gave the Taylor expansion for the PIG probability function, based on the results found in Ong (1995). Generally, Ong's (1998) Taylor expansion method may be applied for any mixed Poisson distribution. Sankaran (1968) used the method of moments estimation when fitting the distribution to European corn beans data and showed that the PIG model fits better than the Hermite distribution. Jørgensen (1987), Stein and Juritz (1988) and Dean, Lawless and Willmot (1989) independently studied the PIG regression model, each using different approaches with their own advantages.

The three-parameter Sichel distribution is an extension of the two-parameter PIG distribution. The Sichel distribution is obtained when the positively-skewed generalized inverse Gaussian (GIG) distribution (Jørgensen, 1982; Nguyen, Chen, Gupta & Dinh, 2003) is used as the mixing distribution in the mixed Poisson formulation. The pmf of the Sichel distribution is given as

$$\Pr(X = k) = \frac{(1-\theta)^{\frac{\gamma}{2}} \left(\frac{\alpha\theta}{2} \right)^k}{k! K_\gamma(\alpha\sqrt{1-\theta})} K_{k+\gamma}(\alpha) \qquad (3.2.3)$$

for k = 0, 1, 2, ..., $\alpha > 0$, $0 < \theta < 1$ and $-\infty < \gamma < \infty$ and $K_\nu(z)$ is the modified Bessel function of the third kind with index ν. Using a moment ratio diagram, Stein, Zucchini and Juritz (1987) illustrated the additional flexibility gained by introducing the additional parameter γ in the three-parameter Sichel distribution which is found to be suitable for modelling over dispersed and long-tailed skewed data. The r-th raw moments of the Sichel distribution is defined as $E(X^r) = \left(\frac{\alpha\theta}{2\sqrt{1-\theta}}\right)^r \frac{K_{\lambda+r}(\alpha\sqrt{1-\theta})}{K_\lambda(\alpha\sqrt{1-\theta})}$. Consequently, the mean and variance of the Sichel distribution are respectively $\mu = \frac{\alpha\theta}{2\sqrt{(1-\theta)}} \frac{K_{\gamma+1}\left(\alpha\sqrt{1-\theta}\right)}{K_\gamma\left(\alpha\sqrt{1-\theta}\right)}$ and $\sigma^2 = \frac{(\alpha\theta)^2}{4(1-\theta)} \frac{K_{\gamma+2}(\alpha\sqrt{1-\theta})}{K_\gamma(\alpha\sqrt{1-\theta})} + \mu[1-\mu]$ (Sichel, 1974). In a series of paper, the applications of this distribution have been discussed by various authors: number of diamond of stones in a sample of gravel (Sichel, 1973), word frequency and sentence-length in linguistics (Sichel, 1974; 1975), consumer behaviour (Sichel, 1982b), insurance claims (Willmot, 1986) and distribution of protein abundance in complex mixtures (Ishihama et al., 2005; Koziol et al., 2013). To facilitate statistical inference procedures, authors such as Stein, Zucchini and Juritz (1987) have proposed a reparameterization of the Sichel distribution. Johnson, Kemp and Kotz (2005) have given a summary of the various reparameterizations of the Sichel distribution.

The Poisson-lognormal distribution is introduced by Bulmer (1974) in the study of species abundance. A disadvantage of this distribution is that there is no closed form for its pmf and it is expressed in the form of an integral

$$\Pr(X = k) = \frac{1}{\sqrt{2\pi}\sigma k!} \int_0^\infty e^{-\theta} \theta^{k-1} \exp\left(-\frac{(\log\theta - \mu)^2}{2\sigma^2}\right) d\theta \qquad (3.2.4)$$

for k = 0, 1, 2, ..., and $-\infty < \mu < \infty$, $\sigma > 0$, where

$$f(\theta) = \frac{1}{\sigma\theta\sqrt{2\pi}} \exp\left(-\frac{(\log\theta - \mu)^2}{2\sigma^2}\right)$$

is the probability density function (pdf) of the lognormal distribution. The mean and variance of the Poisson-lognormal distribution are $\alpha\sqrt{\omega}$ and $\alpha\sqrt{\omega}[\alpha\sqrt{\omega}(\omega - 1) + 1]$, respectively, where $\alpha = e^{\mu}$ and $\omega = \exp(\sigma^2)$. Although it is suitable for fitting long-tailed data, such as in bibliometry (Stewart, 1994), microbial contaminant in food safety risk assessments (Williams & Ebel, 2012) and in modelling species abundance (Bulmer, 1974), the intractability of its pmf limited its application.

Sankaran (1970) presented the Poisson-Lindley distribution and showed that this one-parameter distribution gives a satisfactory fit to two over dispersed data sets. The pmf of the Poisson-Lindley distribution is relatively simple:

$$\Pr(X = k) = \frac{\theta^2(\theta + 2 + k)}{(\theta + 1)^{k+3}}$$

for k = 0, 1, 2, ..., θ > 0. Ghitany and al-Mutairi (2009) showed that the method of moments estimators and maximum likelihood estimators for the Poisson-Lindley parameters are equally efficient, consistent and asymptotically normal. Upon the introduction of a generalized Lindley distribution by Zakerzadeh and Dolati (2009), Mahmoudi and Zakerzadeh (2010) used this generalized Lindley distribution as the mixing distribution in constructing a new mixed Poisson distribution. The Poisson-generalized Lindley distribution has pmf

$$\Pr(X = k) = \frac{\Gamma(k + \alpha)}{k!\Gamma(\alpha + 1)} \frac{\theta^{\alpha+1}}{(\theta + 1)^{k+\alpha+1}} \left(\alpha + \frac{k + \alpha}{\theta + 1}\right)$$

for k = 0, 1, 2, ..., $\theta > 0$, $\alpha > 0$. Mahmoudi and Zakerzadeh (2010) examined the characteristics and statistical inference of this distribution and concluded that it gives a good fit compared to the NB and Poisson-Lindley distributions.

In addition to those mentioned above, various other mixed Poisson distributions have been proposed in the literature, for example, the Poisson-Beta (Holla & Bhattacharya, 1965), Poisson rectangular (Bhattacharya & Holla, 1965) and Poisson-Normal or Hermite distribution (Kemp & Kemp, 1966). Gupta and Ong (2005), Karlis and Xekalaki (2005) and Nikoloulopoulos and Karlis (2008) have given a comprehensive survey on mixed Poisson distributions.

Many properties of a mixed Poisson distribution can be inferred from its mixing distribution. An important result by Holgate (1970) states that a mixed Poisson distribution is unimodal if its mixing distribution is unimodal. Feller (1943) has shown that mixed Poisson distributions are identifiable, i.e. a mixed Poisson distribution and its mixing distribution possess a one-to-one relationship. On the moments of the distributions, the factorial moments of the mixed Poisson distribution are equal to the mixing distribution's moments about the origin. As a result, the r-th moment about the origin of the mixed Poisson distribution can be obtained as

$$E(X^r) = \sum_{j=1}^{r} S(r, j) E(\theta^j), \; r = 1, 2, ...,$$

where $S(n, k)$ denotes the Stirling numbers of the second kind and $E(\theta^j)$ is the j-th moment of the mixing distribution about the origin. Another important result is by Grandell (1997), stating that a mixed Poisson with mixing distribution $f_1(\theta)$ converges to another mixed Poisson with mixing distribution $f_2(\theta)$ if and only if $f_1(\theta)$ converges to $f_2(\theta)$. This implies that the limiting case of a mixed Poisson distribution is "uniquely determined by a limiting case of the mixing distribution" (Karlis & Xekalaki, 2005, p.42). For predicting future outcomes, the posterior expectation of the random variable θ is given by $E(\theta^r \mid X = k) = \frac{\Pr(X = k + r)}{\Pr(X = k)}(k+1)...(k+r)$ (Willmot & Sundt, 1989).

3.3 The Generalized Sichel Distribution

Let Y be a random variable with support on nonnegative real numbers. Then the pdf of the extended generalized inverse Gaussian (EGIG) distribution is given by

$$f(y) = \frac{1}{\frac{2}{\delta}\left(\frac{b}{a}\right)^{\frac{\lambda}{2\delta}} K_{\frac{\lambda}{\delta}}\left(2\sqrt{ab}\right)} y^{\lambda-1} \exp\left(-ay^{\delta} - by^{-\delta}\right), \quad y > 0, \qquad (3.3.1)$$

where $K_{\nu}(z)$ is the modified Bessel function of the third kind with index ν. Here we follow the notation adopted by Gupta and Viles (2011). We adopt a similar domain of variation for the parameters to that given by Jørgensen (1982), that is $\lambda \in \Re$, $(a,b,\delta) \in \Omega_{\lambda}$, where

$$\Omega_{\lambda} = \begin{cases} (a,b,\delta) : a > 0, b \geq 0, \delta > 0 \text{ iff } \lambda > 0 \\ (a,b,\delta) : a > 0, b > 0, \delta > 0 \text{ iff } \lambda = 0 \\ (a,b,\delta) : a \geq 0, b \geq 0, \delta > 0 \text{ iff } \lambda < 0 \end{cases}.$$

When $\delta = 1$, the EGIG model reduces to the generalized inverse Gaussian (GIG) model which has been studied in detail by Jørgensen (1982). Other special and limiting cases of (3.3.1) include the inverse Gaussian distribution ($\delta = 1, \lambda = -0.5$), the gamma distribution ($\delta = 1$, b = 0 and $\lambda > 0$), the Weibull distribution and the exponential distribution.

Definition 3.1 *(Generalized Sichel distribution)* Suppose X is a discrete random variable and $X|\Theta$ ~ Poisson(θ), where Θ is a nonnegative real valued random variable with pdf $f(\theta)$ given by (3.3.1). Then X has the generalized Sichel distribution with pmf given by

$$P(X=k)=\frac{1}{\left(\frac{2}{\delta}\right)\left(\frac{b}{a}\right)^{\frac{\lambda}{2\delta}}K_{\frac{\lambda}{\delta}}(2\sqrt{ab})}\int_0^\infty \frac{e^{-\theta}\theta^k}{k!}\theta^{\lambda-1}\exp(-a\theta^\delta-b\theta^{-\delta})d\theta, \quad (3.3.2)$$

which can be written as

$$P(X=k)=\left(\frac{1}{k!}\right)\frac{\sum_{j=0}^{\infty}\frac{(-1)^j}{j!}\left(\frac{b}{a}\right)^{(j+k)/2\delta}K_{(j+k+\lambda)/\delta}(2\sqrt{ab})}{K_{\lambda/\delta}(2\sqrt{ab})}. \quad (3.3.3)$$

The derivation of expression (3.3.3) is given in Appendix A. The probability generating function (pgf) of the generalized Sichel distribution is given by $G(z)=c\int_0^\infty e^{\theta(z-1)}\theta^{\lambda-1}\exp(-a\theta^\delta-b\theta^{-\delta})d\theta$, yielding

$$G(z)=\frac{\sum_{j=0}^{\infty}\frac{(z-1)^j}{j!}\left(\frac{b}{a}\right)^{j/2\delta}K_{(j+\lambda)/\delta}(2\sqrt{ab})}{K_{\lambda/\delta}(2\sqrt{ab})}, \ |z|\le 1. \quad (3.3.4)$$

From (3.3.4), we obtain the probability when $k = 0$ as

$$\Pr(X=0)=\frac{1}{K_{\lambda/\delta}(2\sqrt{ab})}\sum_{j=0}^{\infty}\frac{(-1)^j}{j!}\left(\frac{b}{a}\right)^{j/2\delta}K_{(j+\lambda)/\delta}(2\sqrt{ab}).$$

The generalized Sichel distribution has mean $\mu = \dfrac{\left(\dfrac{b}{a}\right)^{1/2\delta} K_{(1+\lambda)/\delta}(2\sqrt{ab})}{K_{\lambda/\delta}(2\sqrt{ab})}$ and variance

$\sigma^2 = \dfrac{\left(\dfrac{b}{a}\right)^{1/\delta}}{K_{\lambda/\delta}(2\sqrt{ab})}\left[K_{(2+\lambda)/\delta}(2\sqrt{ab}) - \dfrac{\left(K_{(1+\lambda)/\delta}(2\sqrt{ab})\right)^2}{K_{\lambda/\delta}(2\sqrt{ab})}\right] + \mu$. Consequently, an expression for the index of dispersion can be written as $ID_X = 1 + \left(\dfrac{b}{a}\right)^{1/2\delta}\left[\dfrac{K_{(2+\lambda)/\delta}(2\sqrt{ab})}{K_{(1+\lambda)/\delta}(2\sqrt{ab})}\right] - \mu$.

The special case $\delta = 1$ gives rise to the Sichel distribution. Furthermore, when $\delta = 1$ and $\lambda = -\dfrac{1}{2}$, we obtain the PIG distribution. The Poisson-Gamma (or negative binomial) distribution is obtained from (3.3.2) when $\delta = 1$, b = 0 and $\lambda > 0$. These special cases are derived based on the probability structure of the EGIG model.

The extra parameter δ adds flexibility to the shape of the count distribution. The effect of varying the parameter δ on the index of dispersion ID_X is illustrated in Figure 3.1 in the next page.

3.4 The Poisson-exponentiated Weibull Distribution

Let Y be a random variable having the exponentiated Weibull distribution. The pdf of the exponentiated Weibull distribution is given by Mudholkar et al. (1995) as

$$f(y) = \alpha\eta\beta^{-\alpha} y^{\alpha-1} \exp\left(-\frac{y}{\beta}\right)^{\alpha}\left(1 - \exp\left(-\frac{y}{\beta}\right)^{\alpha}\right)^{\eta-1} \qquad (3.4.1)$$

for y > 0 and α, β, η > 0 with Laplace transform (Choudhury, 2005)

$$L = \eta \sum_{r=0}^{\infty} \frac{(-s\beta)^r \, \Gamma\left(\frac{r}{\alpha}+1\right)}{r!} \left[1 + \sum_{i=1}^{\infty} a_i \left\{ (i+1)^{-(r/\alpha+1)} \right\} \right], \tag{3.4.2}$$

where $a_i = \dfrac{\eta^1(\eta^1-1)\cdots(\eta^1-\overline{i-1})}{i!}(-1)^i$, i = 1, 2, 3, ..., and $\theta^1 = \theta - 1$.

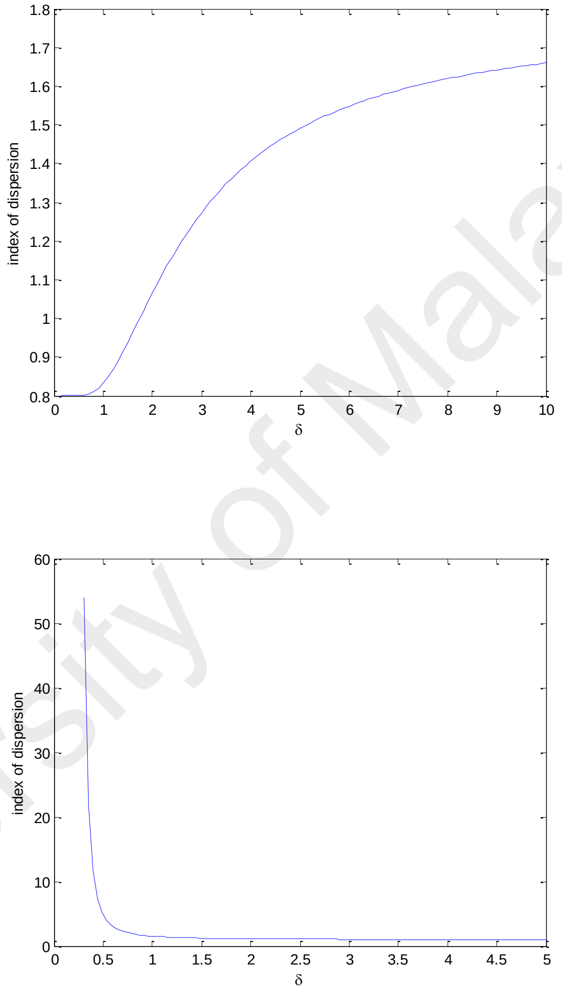


Figure 3.1: Plot of index of dispersion versus δ when (top) *a* = *b* = 0.95 and λ = -0.5, (bottom) *a* = 1.0, *b* = 0.1, λ = -5.

The exponentiated Weibull distribution allows for a non-monotonic hazard function and is introduced as an extension to the Weibull distribution in reliability studies and survival analysis. The parameters α and η are known as the shape parameters whereas β is known as the scale parameter. The Weibull distribution is a special case of this distribution when η = 1. When α = 1 and α = 2, we obtain the exponentiated exponential distribution and Burr Type X distribution, respectively. Furthermore, when both η and α take the value 1, the exponentiated Weibull distribution reduces to the exponential distribution. The *r*-th moment about the origin of the exponentiated Weibull distribution exists in closed form if η is a positive integer.

Definition 3.2 (*Poisson-exponentiated Weibull distribution*) Suppose *X* is a discrete random variable and $X \mid \Theta$ ~ Poisson (θ), where Θ is a nonnegative real valued random variable with pdf $f(\theta)$ given by (3.4.1). Then the probability mass function (pmf) of *X* is given by

$$\Pr(X = k) = \frac{\alpha\eta\beta^{-\alpha}}{k!}\int_0^\infty e^{-\theta}\theta^{k+\alpha-1}\exp\left(-\left(\frac{\theta}{\beta}\right)^{\alpha}\right)\left(1-\exp\left(-\left(\frac{\theta}{\beta}\right)^{\alpha}\right)\right)^{\eta-1} d\theta,\ k = 0, 1, 2, \ldots . \quad (3.4.3)$$

Since the probability generating function (pgf) of a mixed Poisson distribution $G(z)$ is the moment generating function of the mixing distribution evaluated at (*z* - 1) (Karlis & Xekalaki, 2005), we obtain the pgf of the Poisson-exponentiated Weibull distribution as

$$G(z) = \eta\sum_{r=0}^{\infty}\frac{((s-1)\beta)^r\,\Gamma\left(\frac{r}{\alpha}+1\right)}{r!}\left[1+\sum_{i=1}^{\infty}a_i\left\{(i+1)^{-(r/\alpha+1)}\right\}\right].$$

It is straightforward that the probability when *k* = 0 is given as

$$\Pr(X = 0) = \eta\sum_{r=0}^{\infty}\frac{(-\beta)^r\,\Gamma\left(\frac{r}{\alpha}+1\right)}{r!}\left[1+\sum_{i=1}^{\infty}a_i\left\{(i+1)^{-(r/\alpha+1)}\right\}\right].$$

Following the result by Holgate (1970), the Poisson-exponentiated Weibull distribution is unimodal when $\alpha\eta > 1$.

3.4.1 The Poisson-Weibull Distribution

A special case when $\eta = 1$ gives rise to the Poisson-Weibull distribution which has been discussed by Cheng et al. (2013) in the context of a generalized linear model. The Weibull distribution is of special importance in the field of reliability studies, partly due to its ability to model devices with constant, increasing and decreasing failure rates.

The pgf of the Poisson-Weibull distribution can be written as

$$G(z) = \sum_{n=0}^{\infty} \frac{(z-1)^n \beta^n}{n!} \Gamma\left(1 + \frac{n}{\alpha}\right). \tag{3.4.4}$$

The mean and variance of the Poisson-Weibull distribution can be derived from the Weibull distribution. Since the Weibull distribution has mean and variance $\beta\Gamma\left(\alpha^{-1}+1\right)$ and $\beta^2\left\{\Gamma\left(2\alpha^{-1}+1\right)-\left[\Gamma\left(\alpha^{-1}+1\right)\right]^2\right\}$, respectively, the formula to obtain the mean μ, variance σ^2 and index of dispersion ID_X of the Poisson-Weibull distribution are as follows:

$$\mu = \beta\Gamma\left(\alpha^{-1}+1\right), \tag{3.4.5}$$

$$\sigma^2 = \mu + \beta^2\Gamma\left(2\alpha^{-1}+1\right) - \mu^2 \text{, and} \tag{3.4.6}$$

$$ID_X = 1 + \beta \frac{\Gamma(2\alpha^{-1}+1)}{\Gamma(\alpha^{-1}+1)} - \mu. \tag{3.4.7}$$

3.5 Shape of the Distributions

The generalized Sichel distribution is a flexible model which is able to model data with zero-inflation, over dispersed and long-tailed data. Three examples of the generalized Sichel pmf plots are given in Figure 3.2 to illustrate the versatility of the shape of the distribution.

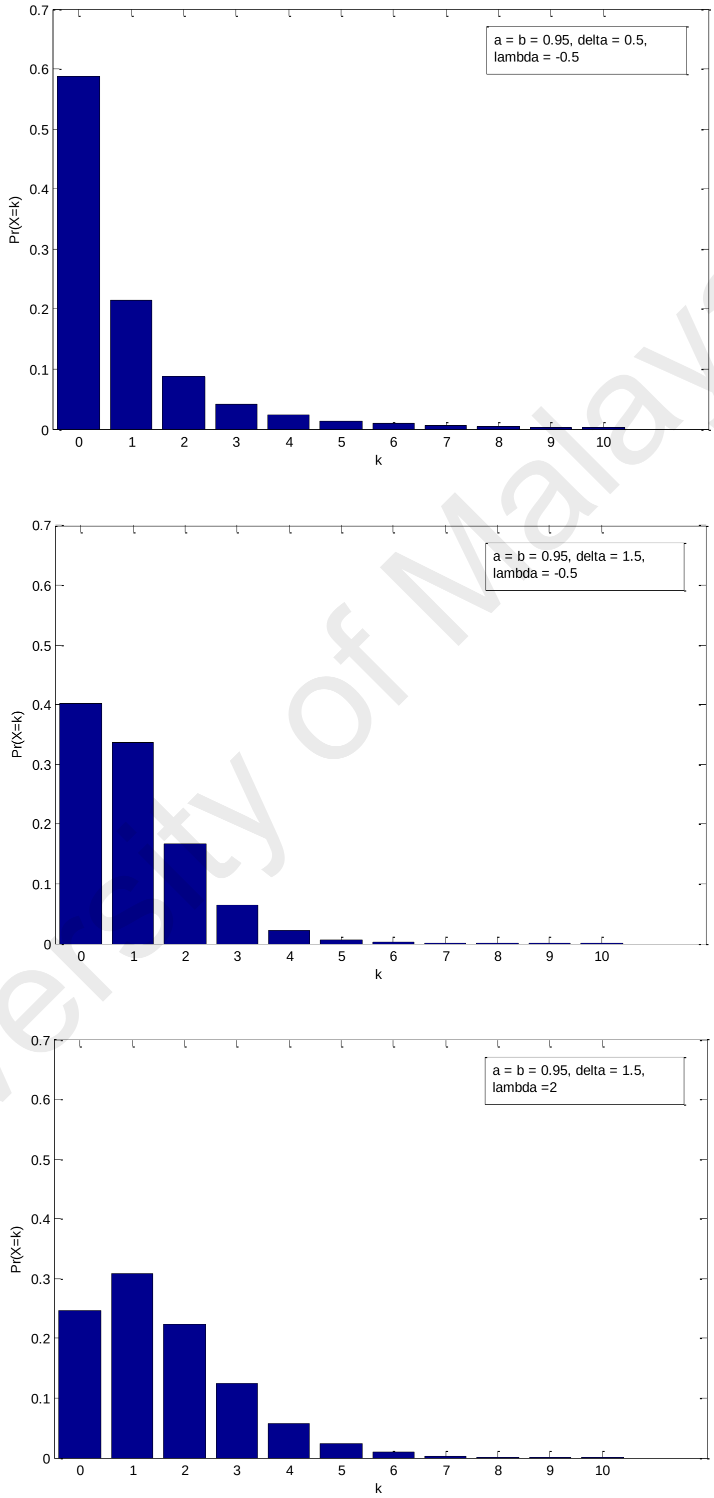
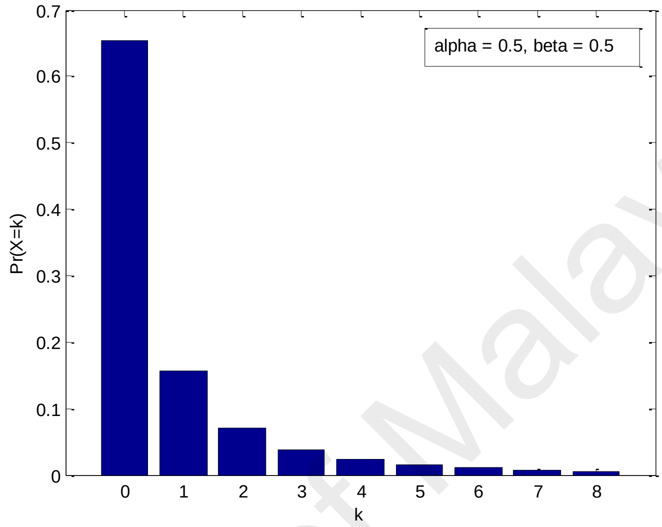


Figure 3.2: Probability mass function plots of the generalized Sichel distribution

To illustrate features such as the position of the mode, skewness and tail length, examples of the Poisson-exponentiated Weibull probability plots are given in Figure 3.3. It is observed that the distribution is able to model data with high zero counts, right-skewed or almost symmetric data.



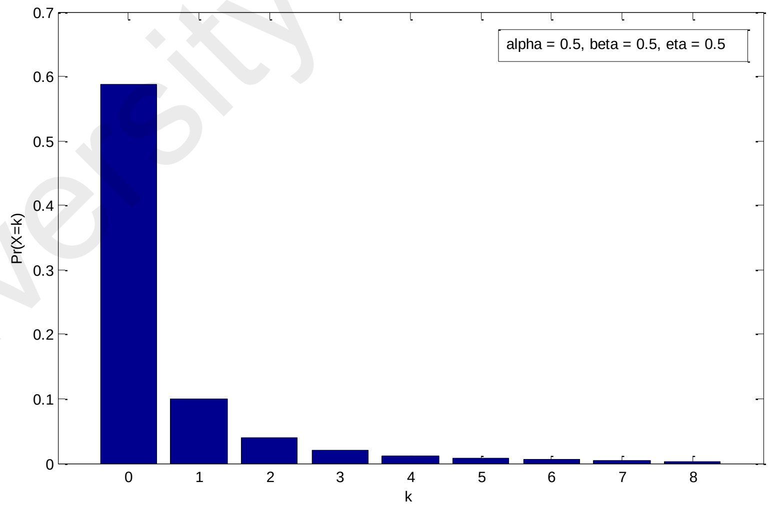
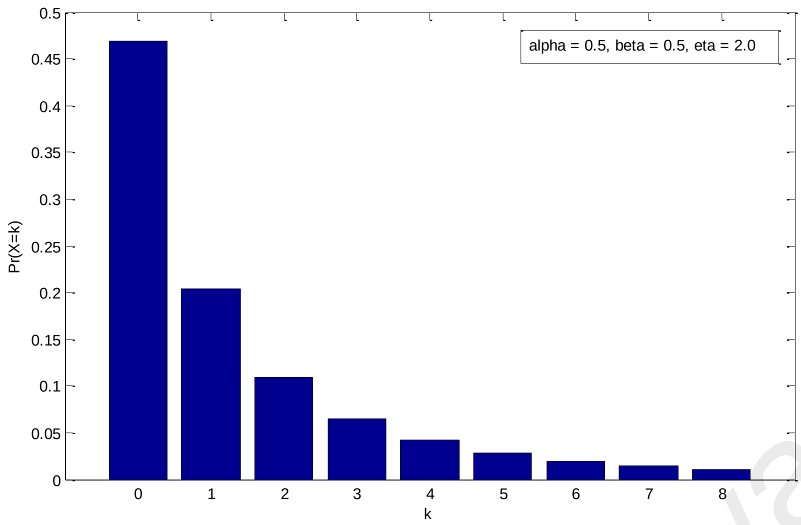
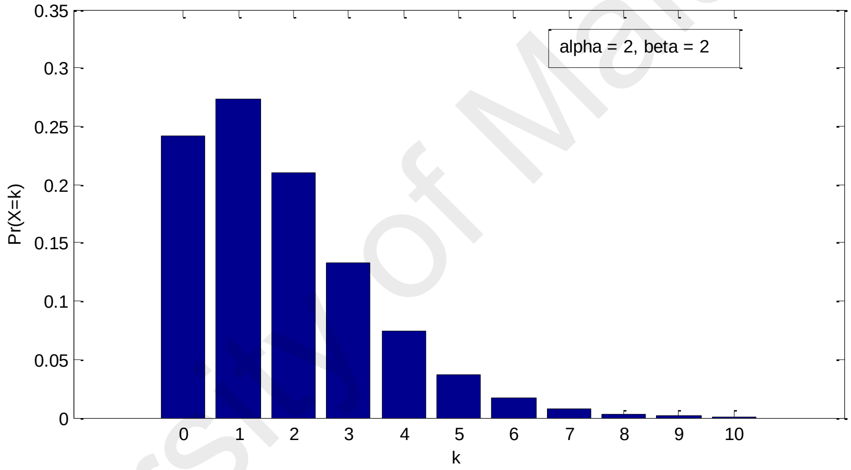


Figure 3.3: Probability mass function plots of the Poisson-exponentiated Weibull distribution





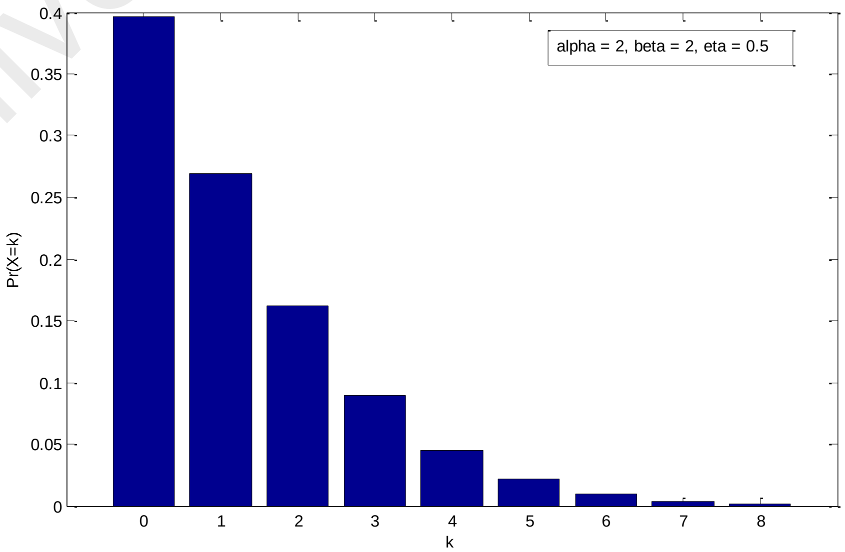
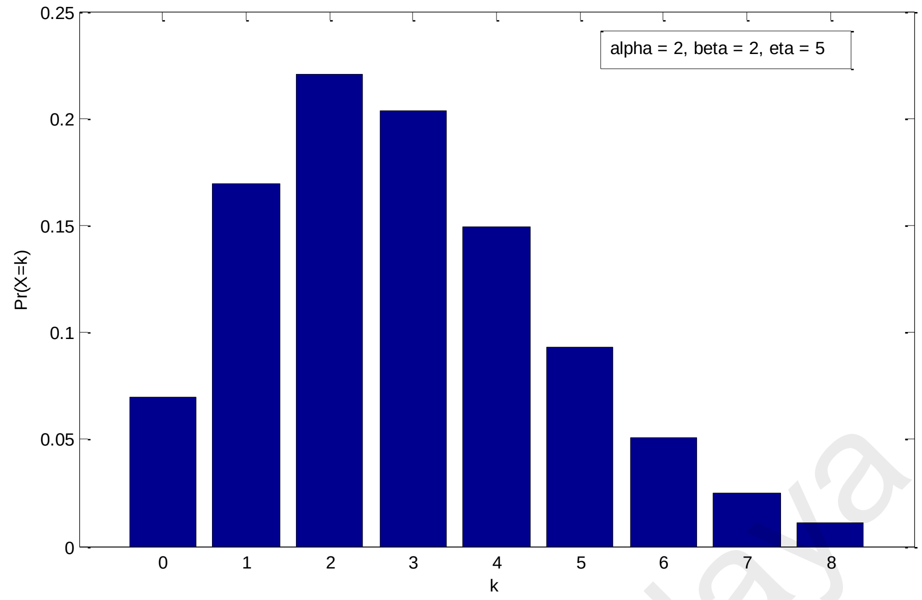


Figure 3.3, continued



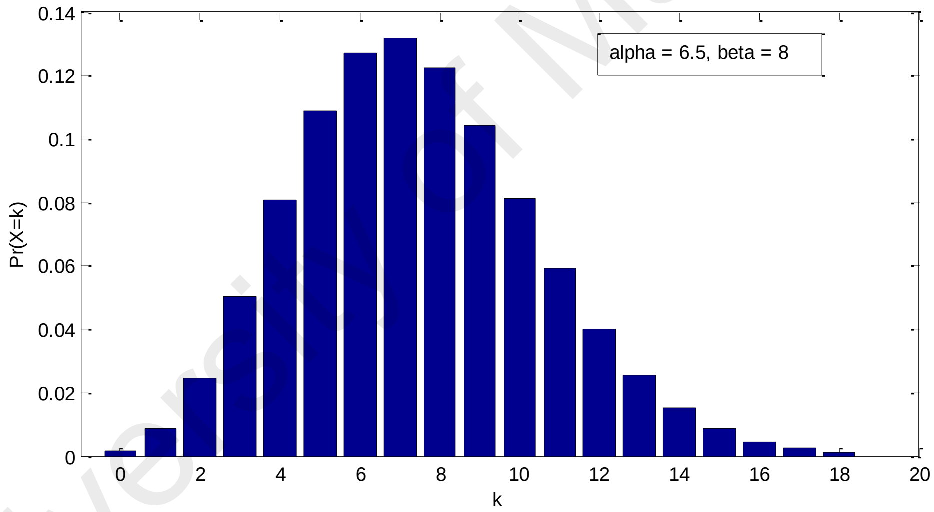


Figure 3.3, continued

The Poisson-Weibull distribution is a special case of the Poisson-exponentiated Weibull family when $\eta=1$. A numerical check reveals that when $\alpha < 1$, the Poisson-Weibull distribution has a high mode at k = 0 when β is small. As the value of β increases, the probabilities shifts away from k = 0 but the mode retains. On the other hand, when $\alpha > 1$, the mode shifts away from k = 0 as the value of β increases and the distribution changes from being right-skewed to almost symmetrical. The larger the value of α, the further the mode shifts away from $k = 0$ when β increases.

In general, increasing the value of η shifts the probabilities from the zero counts to the non-zero counts thus the parameter η affects the lower tail of the probability distribution.

3.6 Some Characteristics of Mixed Poisson Distributions

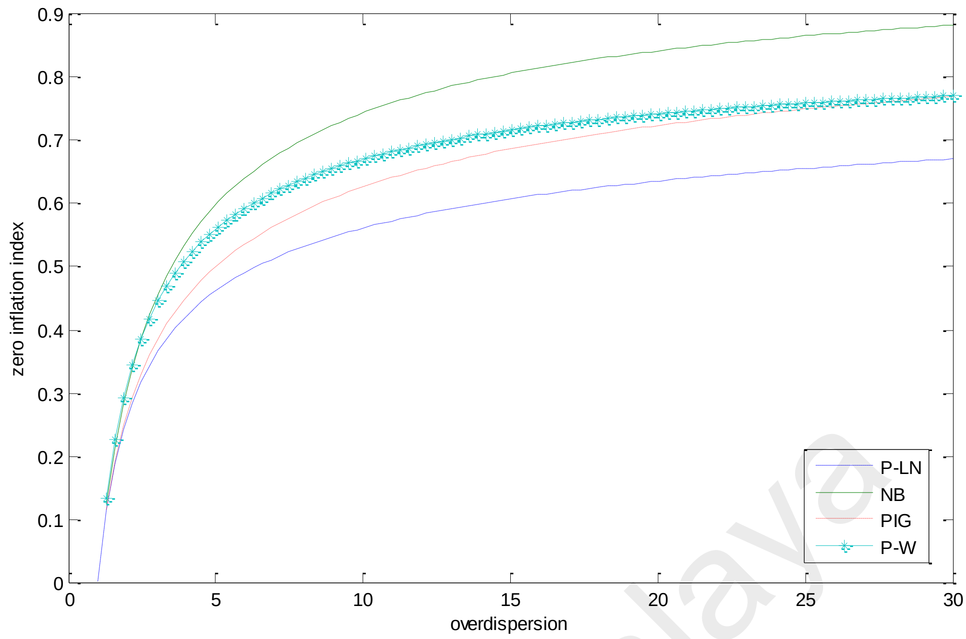
We examine the shape of the mixed Poisson distributions discussed in the preceding sections in terms of the zero-inflation index and the third central moment inflation index as defined by Puig and Valero (2006). For the generalized Sichel and Poisson-exponentiated Weibull distributions, both the zero-inflation and the central moment indices are dependent on the mean and they are obtained using numerical computation.

3.6.1 Zero-inflation Index

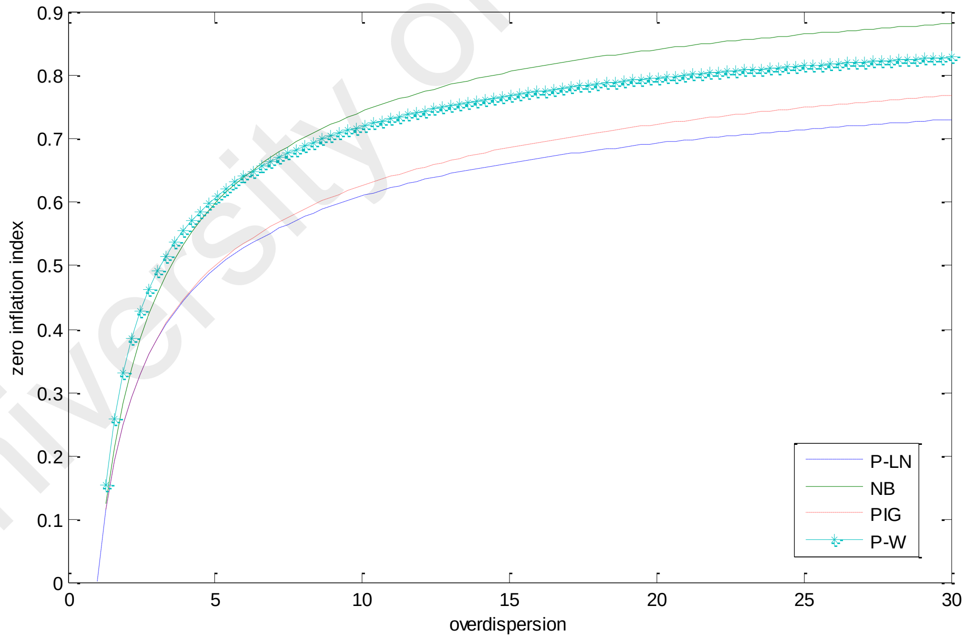
The zero-inflation index of a nonnegative integer random variable X with mean μ and proportion of zeros p_0 is defined as $zi = 1 + \log(p_0)/\mu$ (Puig & Valero, 2006). The Poisson random variable has a zero-inflation index of 0, and a zero-inflated random variable will have a positive zero-inflation index. It is known that any mixed Poisson random variable is zero-inflated thus it is of interest to know the amount of zero-inflation.

The zero-inflation index for NB and PIG are independent of the mean of the distribution and it is expressed as $1 + \log(ID)/(1 - ID)$ and $(ID - \sqrt{2(ID) - 1})/(ID - 1)$, respectively, where ID is the index of dispersion (Nikoloulopoulos & Karlis, 2008). In Figure 3.4, we plot the zero-inflation index versus index of dispersion for some biparametric mixed Poisson distributions: Poisson-lognormal (P-LN), negative binomial (NB), Poisson-inverse Gaussian (PIG) and Poisson-Weibull (P-W) distributions. The zero-inflation index of the P-LN and P-W distributions are computed numerically for a fixed mean. By comparing the three graphs in Figure 3.4, we see that as the mean increases, the zero-inflation index of the P-W distribution increases. In terms of zero-inflation, the P-W distribution is closer to the NB and PIG distribution than to the P-LN distribution. When both mean and over dispersion is small, the P-W distribution is very close to the NB distribution. On the other hand, when mean is small and over dispersion is large, the P-W distribution becomes closer to the PIG distribution which has a lower zero-inflation index. When the mean is large with small over dispersion, the P-W distribution has the highest zero-inflation index amongst the two-parameter mixed Poisson distributions compared.

We plot the graph of the zero-inflation index for the generalized Sichel (GS) and its related mixed Poisson distributions in Figure 3.5. The zero-inflation index for the GS and Sichel distributions are both computed numerically. In Figure 3.5, the mean for the Sichel distribution is fixed at 2.5. The graph for the Sichel distribution does not change significantly as its mean increases hence it is not included here. We consider the cases when mean = 2.5, 5 and 15 for the GS distribution, representing different sizes of the mean from small to large. The zero-inflation index of the GS distribution increases with the value of its mean. When the mean is large, for example mean = 15, the zero-inflation index of the GS distribution is always larger than the PIG and Sichel distributions. When over dispersion is very small, all three distributions are similar. However, as the mean gets larger, the GS distribution has a higher zero-inflation index than the NB distribution. Compared to the NB distribution, the GS distribution is able to model high zero counts even when presence of over dispersion is small and mean is large.

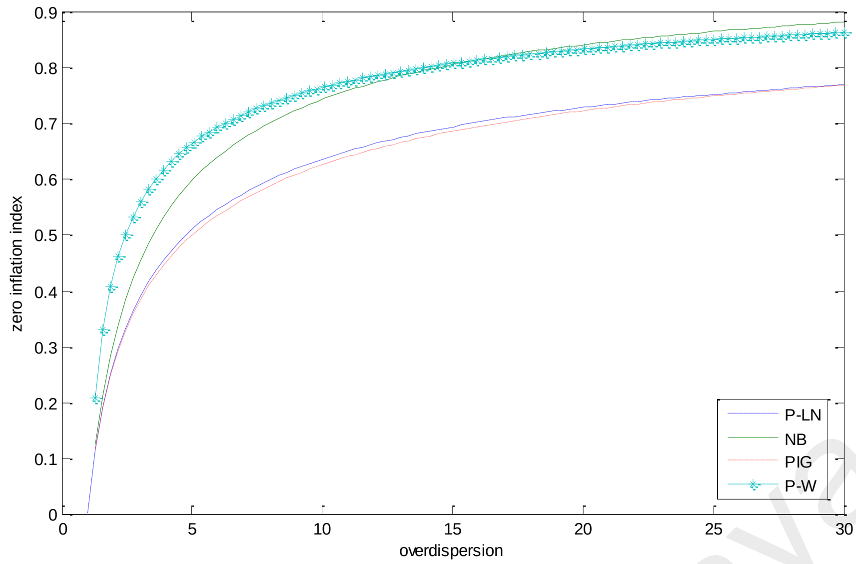


(a) Mean of the Poisson-lognormal and Poisson-Weibull is 1.5



(b) Mean of the Poisson-lognormal and Poisson-Weibull is 5

Figure 3.4: Zero-inflation index versus index of dispersion for the Poisson-Weibull and some biparametric mixed Poisson distributions



(c) Mean of the Poisson-lognormal and Poisson-Weibull is 15

Figure 3.4, continued

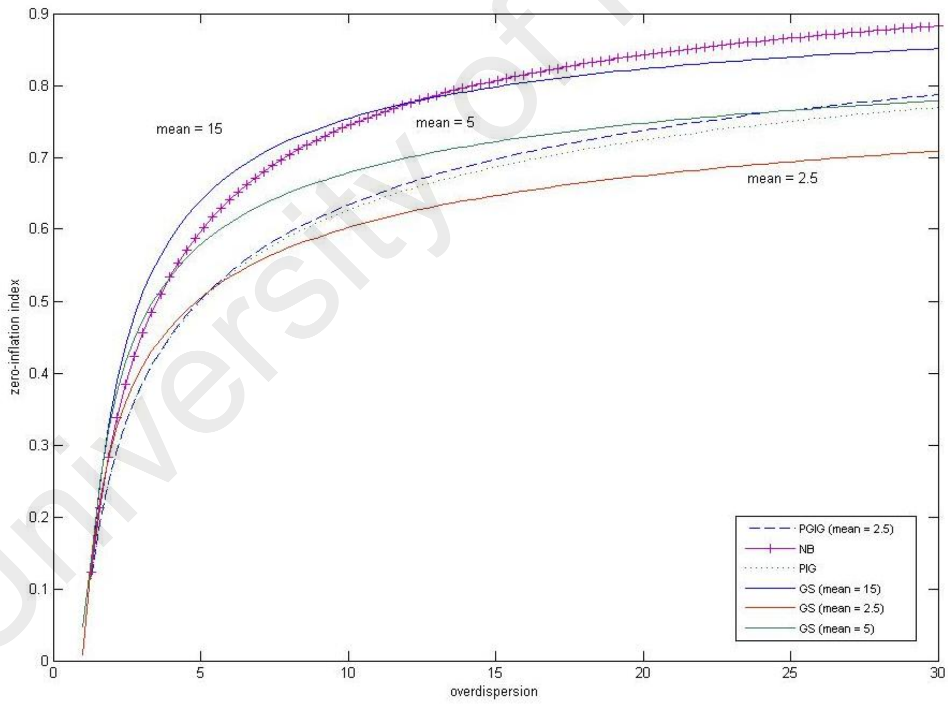


Figure 3.5: Zero-inflation index versus index of dispersion for the generalized Sichel and some related mixed Poisson distributions

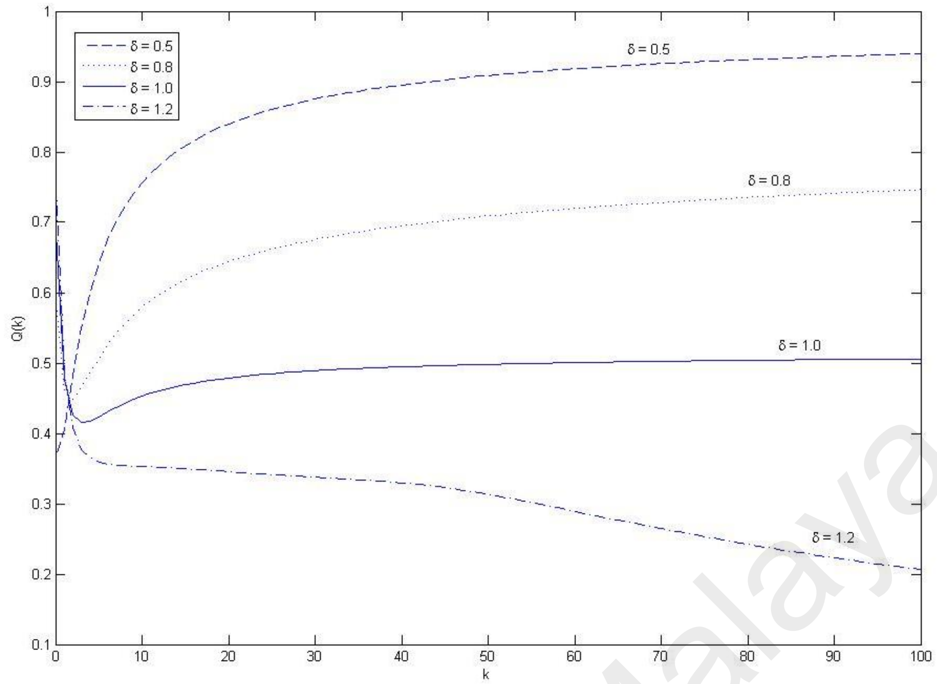
3.6.2 Discriminant Ratio

Ong and Muthaloo (1995) discussed the role of the discriminant ratio which is defined as $Q(k) = \frac{P(X = k+1)}{P(X = k)}$, for k = 0, 1, 2, 3, …, in determining the flexibility of the distributions which they proposed for long-tailed data. The ratio has a limiting value of $Q(k) \to 1$ for long-tailed distributions. Figure 3.6 gives the graphs of $Q(k)$ versus k for several values of the parameter δ of the generalized Sichel distribution, holding other parameters fixed.

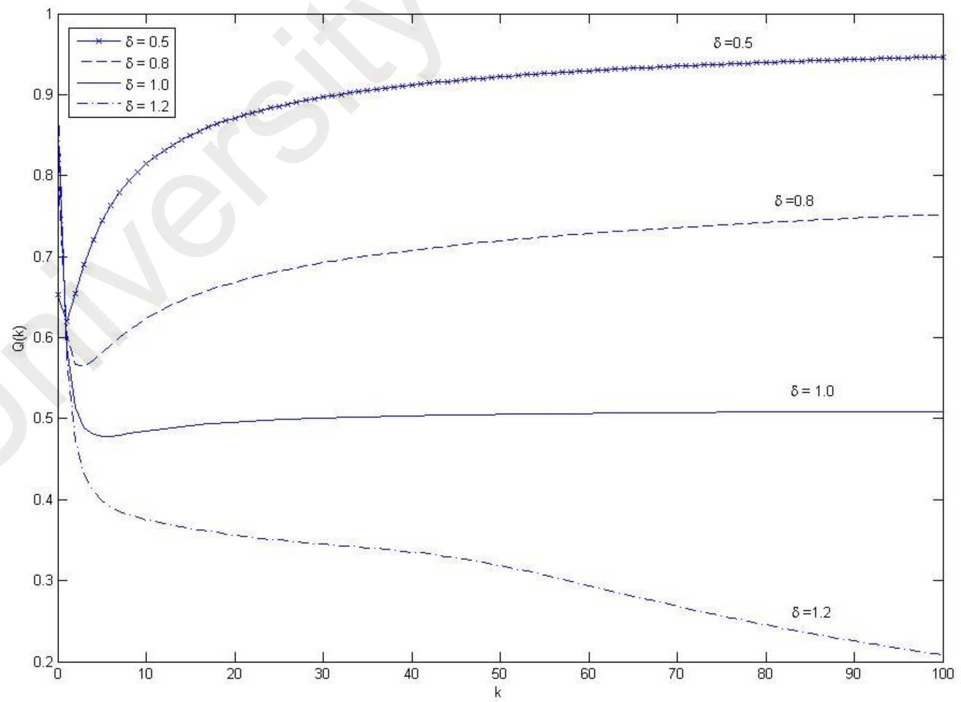
In Figure 3.6(a), we compare the graph of $Q(k)$ versus k for the PIG distribution (δ = 1, λ = -0.5) and the generalized Sichel distribution. By varying the value of δ, the discriminant ratio varies considerably especially at large values of k. The difference is most prominent for k larger than 10.

From the graphs in Figure 3.6(b), we note that when $\delta < 1$, the generalized Sichel distribution has a longer tail compared to the Sichel distribution. The trend is similar to that in Figure 3.6(a). As such, we may conclude that the parameter δ adds flexibility to the generalized Sichel distribution, enabling the distribution to model data with a very long tail.

In the exponentiated Weibull distribution, the parameter η modifies the lower (or left) tail of the distribution. The graphs of $Q(k)$ versus k for the Poisson-exponentiated Weibull distributions in Figure 3.7 show that the parameter η modifies the rate of change of frequencies near the origin. Figure 3.7(a) indicates long-tailed distributions since $Q(k) \to 1$ as $k \to \infty$. Inheriting the right tail properties of their respective mixing distributions, the Poisson-exponentiated Weibull distribution and the Poisson-Weibull distribution have the same behaviour at the right-tail.



(a) Discriminant ratio diagrams when *a* = *b* = 0.95 and λ = -0.5 for PIG (δ = 1) and generalized Sichel distributions

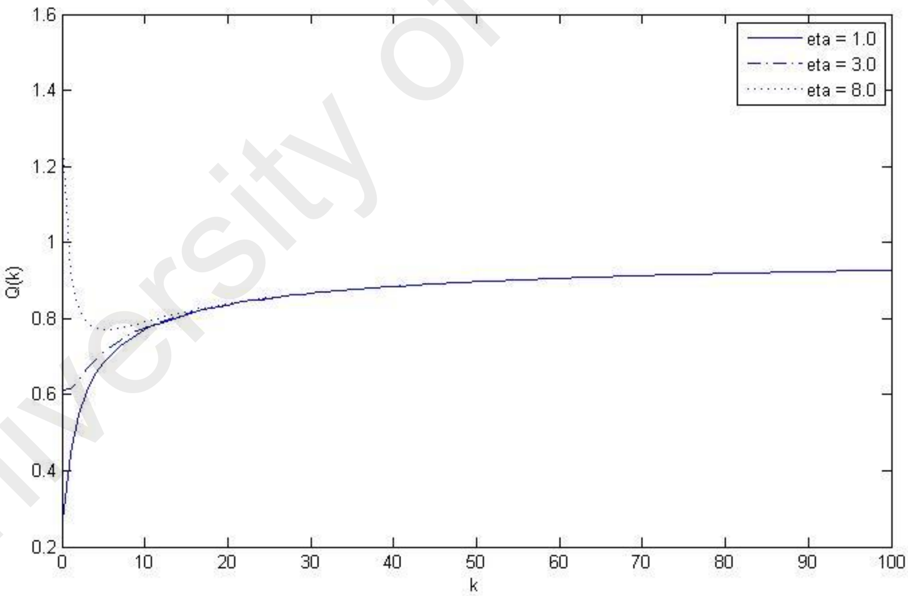


(b) Discriminant ratio diagrams when *a* = *b* = 0.95 and λ = 0.2 for Sichel (δ = 1) and generalized Sichel distributions

Figure 3.6: Discriminant ratio diagrams for generalized Sichel distribution

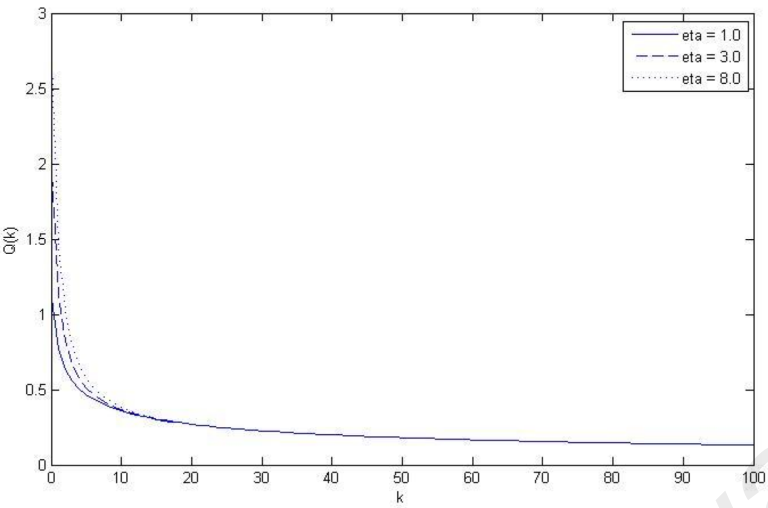
3.6.3 Third Central Moment Inflation Index

The third central moment inflation index of a nonnegative discrete random variable X describes the skewness of the distribution and is defined as $\kappa_3 = \frac{\mu_3}{\mu} - 1$, where μ_3 is the third central moment of X. For discrete distributions, $\mu_3 = E(X^3) - 3E(X^2)E(X) + 2[E(X)]^3$, where $E(X^r)$ is the r-th moment about the origin. The third central moment inflation index takes the value zero for the Poisson distribution. The index for the NB and PIG distributions take the values $2(ID)^2$ - ID - 1 and $3(ID)^2 - 3(ID)$, respectively (Nikoloulopoulos & Karlis, 2008). The third central moment of the PGIG distribution is $E(X^3_{PGIG}) = \left(\frac{\alpha\theta}{2\sqrt{1-\theta}}\right)^3 \frac{K_{\lambda+3}(\alpha\sqrt{1-\theta})}{K_\lambda(\alpha\sqrt{1-\theta})}$.

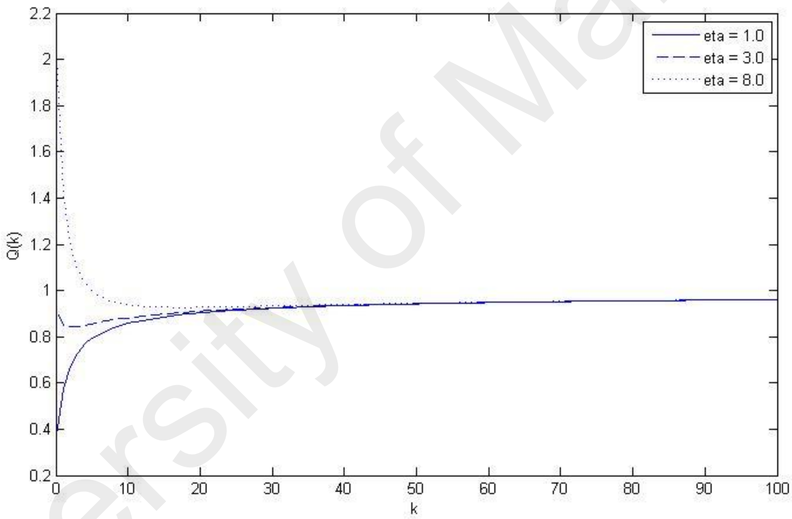


(a) α = β = 0.5

Figure 3.7: Discriminant ratio diagrams for exponentiated Weibull distributions



(b) $\alpha = \beta = 2$

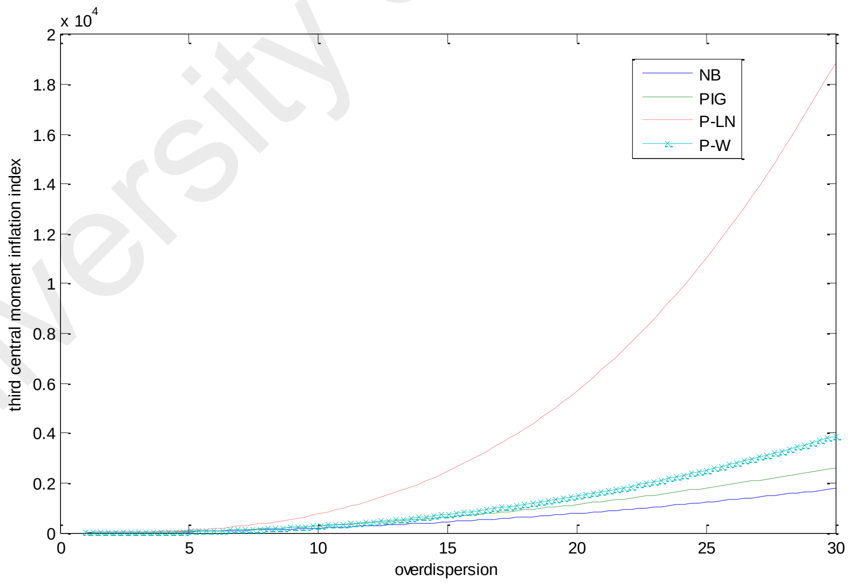


(c) $\alpha = 0.5, \beta = 2$

Figure 3.7, continued

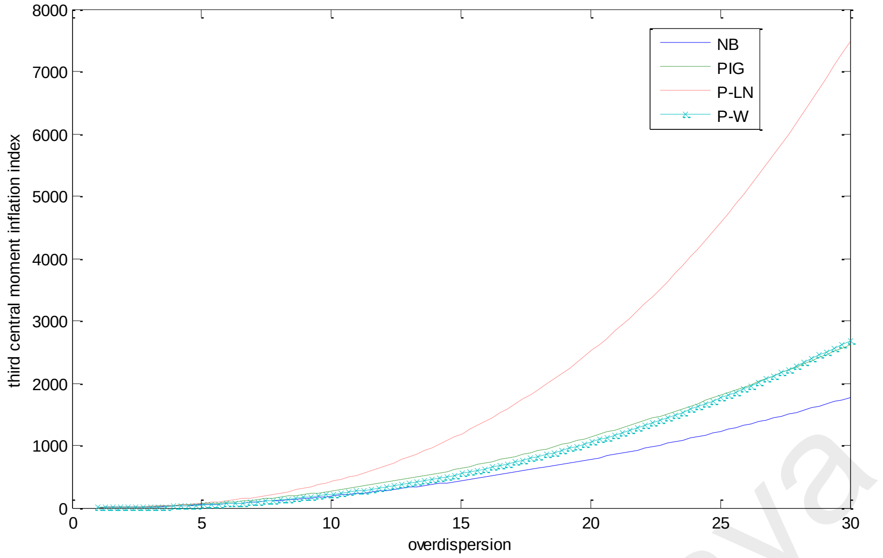
The graphs of the third central moment inflation index versus over dispersion for the Poisson-Weibull and some related biparametric mixed Poisson distributions are given in Figure 3.8. The plot is given for three different values of the Poisson-Weibull and Poisson-lognormal means ranging from small to large in Figure 3.8(a)-(c). In all cases, the index for the Poisson-Weibull distribution is always larger than that for the negative binomial distribution. When presence of over dispersion is small, all four distributions are similar. For small values of the mean, the skewness for the Poisson-Weibull distribution is higher than the PIG distribution. The Poisson-lognormal distribution invariably has the highest skewness index.

In Figure 3.9, we plot the third central moment inflation index versus over dispersion for the NB, PIG, PGIG and generalized Sichel (GS) distributions. For the PGIG and generalized Sichel distributions, the graphs are given for three different values of the mean, i.e. 3, 5, 15. The coefficient of skewness is positive for all four distributions. For small over dispersion, all of the distributions are similar to each other. As the index of dispersion increases, the coefficient of skewness of all the distributions increases. For the generalized Sichel distribution, as the mean increases, the coefficient of skewness decreases. Moreover, the skewness of the generalized Sichel distribution moves closer to the NB distribution as the mean increases. The generalized Sichel distribution has a higher coefficient of skewness than the Sichel distribution when mean is small.

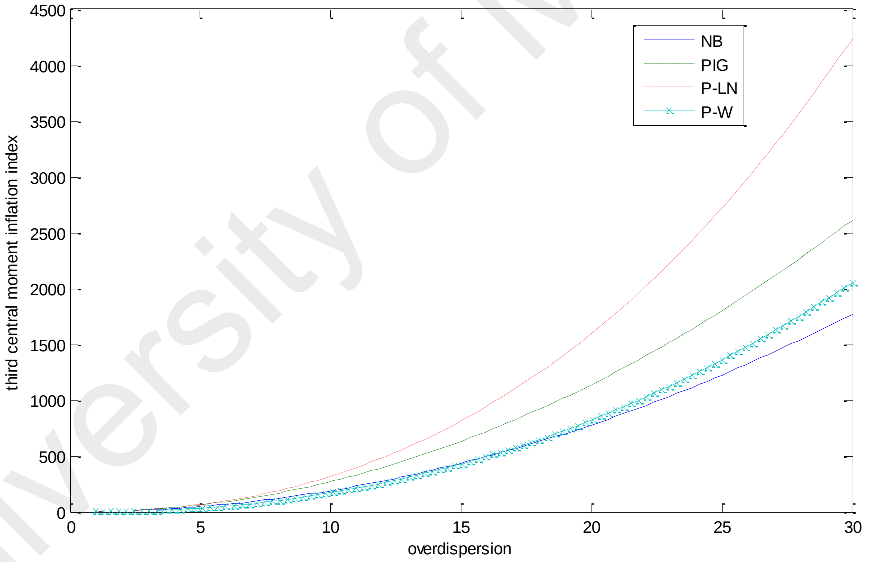


(a) Mean of the Poisson-Weibull and Poisson-lognormal is 1.5

Figure 3.8: Third central moment inflation index versus index of dispersion for Poisson-Weibull and some biparametric mixed Poisson distributions

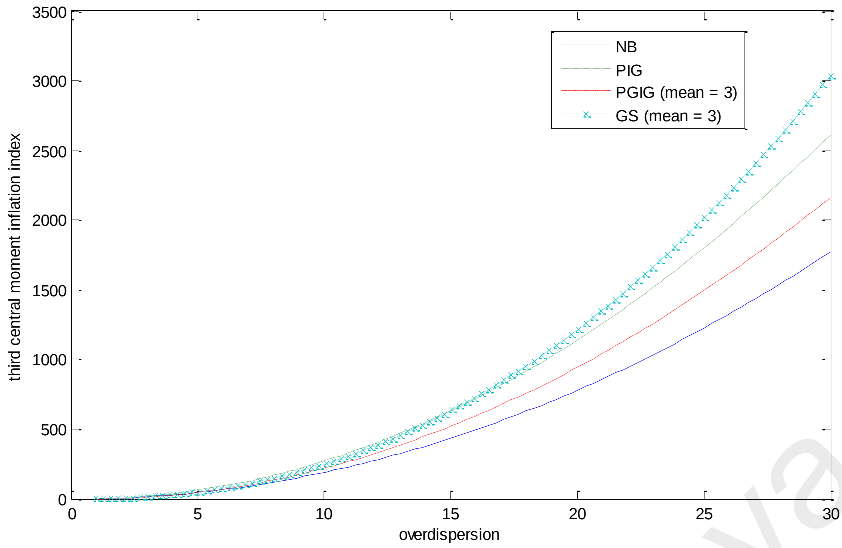


(b) Mean of the Poisson-Weibull and Poisson-lognormal is 5

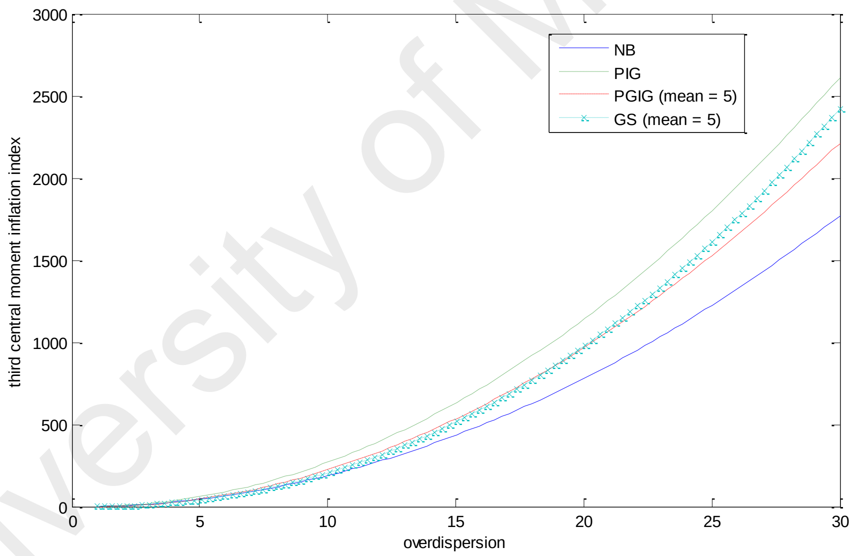


(c) Mean of the Poisson-Weibull and Poisson-lognormal is 15

Figure 3.8, continued

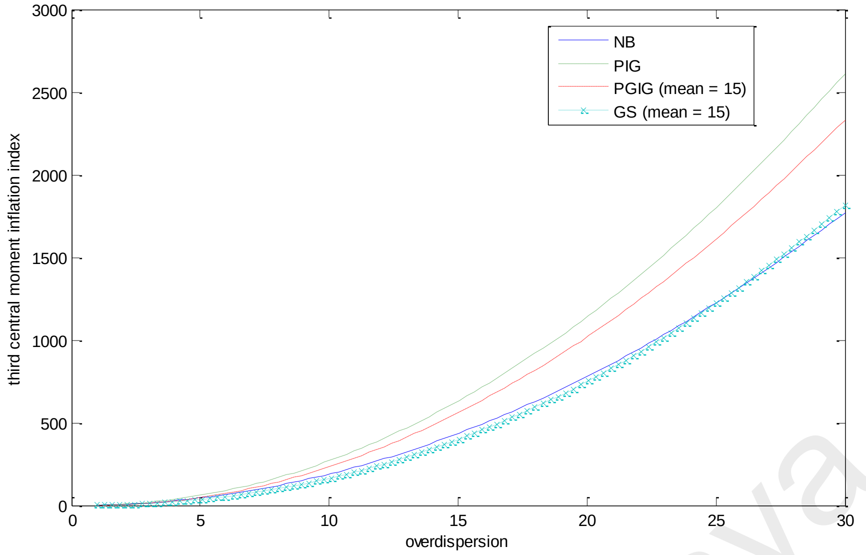


(a) Mean of the PGIG and generalized Sichel is 3



(b) Mean of the PGIG and generalized Sichel is 5

Figure 3.9: Third central moment inflation index versus index of dispersion for generalized Sichel and related mixed Poisson distributions



(c) Mean of the PGIG and generalized Sichel is 15

Figure 3.9, continued

3.7 Conclusion

In this chapter, the generalized Sichel distribution which is an extension of the Sichel (i.e. PGIG) distribution is presented. Based on the characteristics examined in the preceding sections, the generalized Sichel distribution is suitable for modelling over dispersed data with high zero counts and a very long tail. Since the generalized Sichel distribution nests the NB, PIG and PGIG distributions as special cases, this proposed distribution has an advantage of eliminating a piecewise treatment in empirical modelling.

We also examine the Poisson-exponentiated Weibull distribution as a new mixed Poisson distribution suitable for modelling over dispersed data with high zero count and a long tail. A special case of this distribution, the Poisson-Weibull distribution, has a simple closed form for its moments and the distribution is further analysed and compared with other biparametric mixed Poisson distributions.

CHAPTER 4: COMPUTATION AND STATISTICAL INFERENCE FOR MIXED POISSON DISTRIBUTIONS

4.1 Introduction

The pmf of a mixed Poisson distribution is expressed as the integral

$$\Pr(X = k) = \int_0^\infty \frac{e^{-\theta}\theta^k}{k!} f(\theta)d\theta \text{ , } k = 0, 1, 2, ..., \quad (4.1.1)$$

where $f(\theta)$ is the pdf of the mixing distribution. It is favourable to simplify and re-write (4.1.1) in closed form, whenever possible such as in the case of the negative binomial (Poisson-gamma), Sichel or Poisson-Lindley distributions. However, in many cases, a closed form representation cannot be obtained. Consequently, not many mixed Poisson distributions are studied in depth due to the intractability of their probability mass functions. For example, the Poisson-lognormal pmf cannot be written in closed form and thus it has to be evaluated using numerical techniques (Bulmer, 1974; Izsák, 2008) with due consideration for large counts *k*. Cheng, Geedipally and Lord (2013) proposed the Poisson-Weibull distribution for modelling crash count data but noted that the pmf cannot be expressed in closed form. To circumvent the problem, they wrote the model using a hierarchical structure that can be modelled using Bayesian approach in generalized linear models.

In Chapter 3, we have presented two new mixed Poisson distributions, namely the generalized Sichel and Poisson-exponentiated Weibull (P-EW) distributions. Upon studying their properties such as the zero-inflation index and discriminant ratio, these two distributions have been found to be able to model over dispersed discrete count data with high probabilities at zero counts and/or a very long tail. In this chapter, we present a Monte Carlo simulation approach to the computation of P-EW probabilities, which eliminates the need for using numerical methods and specific consideration and analysis for various cases of k, is used. The closed form of the exponentiated Weibull cumulative distribution function (cdf) facilitates the rapid computer generation of exponentiated Weibull Monte Carlo random samples by the method of inverse cdf. As such, the Monte Carlo approach is easy to implement and it is shown to be highly accurate for the P-EW distribution in particular, and in general for any mixed Poisson distribution when the mixing distribution has a closed form cdf. We also study an Expectation-Maximization (EM) type algorithm for parameter estimation, which can eliminate the need to evaluate the integral in the pmf of the Poisson-Weibull distribution.

The remaining of the chapter proceeds as follow. In the next section, a literature review on computation and statistical inference of mixed Poisson distributions is given. After that, we discuss the Monte Carlo technique for computation of the P-EW probabilities. In Section 4.4, we discuss ML estimation for mixed Poisson parameters, particularly for the generalized Sichel distribution and Poisson-Weibull distribution presented in Chapter 3. In the subsequent section, we discuss the likelihood ratio test for mixed Poisson distributions and a score test statistic is derived for the generalized Sichel distribution. The model fit of over dispersed data sets to some mixed Poisson distributions is presented and examined in Section 4.6. Finally, the conclusion is given in the last section.

4.2 Literature Review

Due to the intractability of the pmf of many mixed Poisson distributions, the computation of mixed Poisson probabilities is of interest. Straightforward evaluation of the mixed Poisson integral can be done using numerical integration techniques such as Gaussian quadrature. Willmot (1990) derived a formula to estimate these probabilities and also showed that the right tail of the Poisson mixture has the same behaviour as the mixing distribution. Perline (1998) derived a similar result. Willmot’s (1990) formula is based on the fact that the right tail of a mixed Poisson distribution inherits the behaviour of its mixing distribution. Bulmer (1974) obtained an approximation of the Poisson-lognormal probabilities for large values of k by re-expressing the pmf as an expectation with respect to a gamma density function, and subsequently applying a Taylor series expansion. Probabilities for small k can be obtained using numerical integration, for example, Grundy (1951) has tabled the probability values when k = 0, 1 using an infinite series approximation to the integral. Cassie (1962) has discussed the use of the lognormal distribution instead of the Poisson-lognormal distribution to approximate the mixed Poisson probabilities. While discussing the ML estimation of the Poisson-lognormal distribution, Izsák (2008) made prior modifications to the integral before applying numerical integration for small k, but proposed the saddle point method when k is large. Ong (1995) discussed a three-term recurrence relation for the computation of the full beta model with pmf given as $\Pr(X = k) = \frac{1}{B(p,q)} \int_0^\infty \frac{e^{-\theta}\theta^k}{k!} \frac{b^p \theta^{p-1}}{(1+b\theta)^{p+q}} d\theta$ for k = 0, 1, ..., b, p, q > 0 and $B(p, q)$ is the Beta function. This distribution is obtained by taking the beta distribution of the second kind with a scale parameter $1/b$ as the mixing distribution. However, this method inherits the stability issues from three-term recurrence relations. Karlis and Xekalaki (2005) proposed a method that expresses the probability function of the mixed Poisson distribution as an infinite series involving the moments of the mixing distribution.

Parameter estimation has been a topic of interest in the literature on mixed Poisson distributions. An issue which arises from the intractability of the mixed Poisson's pmf is in ML estimation procedure since analytical ML estimates require solving the first derivative of the log-likelihood function. Sichel (1982a) studied the asymptotic efficiency of his proposed method for the ML estimation of Poisson-inverse Gaussian (PIG) parameters, but this method is only applicable for a subset of the parameter values. Atkinson and Yeh (1982) derived an approximate likelihood method for the inference for one of the parameters in the Sichel distribution and compared it against ML estimation. However, Stein, Zucchini and Juritz (1987) found that this approximate method may return poor estimates. In view of this, they proposed an alternative parameterization for the Sichel pmf and gave an algorithm for computing the ML estimates. With this new parameterization, the ML estimates have the advantage of being asymptotically uncorrelated. Highly correlated ML estimates may lead to misleading inference and numerical instability during the maximization of the likelihood function.

In a general framework on mixed Poisson distributions, Karlis (2005) discussed an alternative approach to solving the ML estimation problem through the application of an Expectation-Maximization (EM) type algorithm. Karlis' (2005) adaptation of the EM algorithm to the problem of parameter estimation in mixed Poisson distributions reduces the problem of maximizing the mixed Poisson's likelihood function to a simpler problem of maximizing the mixing distribution's likelihood function. Therefore, when the ML estimates of the mixing distribution can be obtained analytically, ML estimation of the mixed Poisson parameters via the EM algorithm becomes a very attractive approach. When the mixed Poisson pmf is intractable, a Monte Carlo variant of the EM algorithm (Wei & Tanner, 1990) can eliminate the need to perform numerical integration for the mixed Poisson probabilities.

4.3 Computational Method for Some Mixed Poisson Distributions

The pmf of the generalized Sichel distribution proposed in Chapter 3 can be expressed as an integral or an infinite series. Based on the integral representation, we use a MATLAB numerical integration algorithm quade.m proposed by Sermutlu and

Eyyuboglu (2007) to compute generalized Sichel probabilities. This algorithm is able to handle improper integrals and non-smooth functions.

The evaluation of integral (4.1.1) can also be studied using a Monte Carlo simulation approach. This Monte Carlo simulation technique for computing integrals eliminates the need for numerical evaluation of the probabilities that can cause stability issues in parameter estimation procedures such as ML estimation. When the cdf of the mixing distribution can be written in closed form, this Monte Carlo approach is simple and fast. We apply the formulation of this technique for the Poisson-exponentiated Weibull distribution proposed in Section 3.4. The Monte Carlo technique is briefly described first in general.

Suppose we have a random variable Θ having pdf $f(\theta)$. We would like to evaluate

$$E[g(\Theta)] = \int g(\theta) f(\theta) d\theta \tag{4.3.1}$$

for some function $g(\theta)$. A crude Monte Carlo estimator uses n simulated realizations $\theta_1, \theta_2, ..., \theta_n$ of Θ to approximate the integral by

$$E[g(\Theta)] \approx \frac{1}{n} \sum_{i=1}^{n} g(\theta_i) \tag{4.3.2}$$

where the expectation operator is taken with respect to the density function $f(\theta)$.

As such, in principle the mixed Poisson probabilities given by (4.1.1) may be approximated by (4.3.2):

$$\Pr(X = k) = \int_0^\infty \frac{e^{-\theta}\theta^k}{k!} f(\theta) d\theta \approx \frac{1}{n} \sum_{i=1}^{n} g(\theta_i; k) \tag{4.3.3}$$

where $g(\theta_i;k)=\frac{e^{-\theta_i}\theta_i^{\,k}}{k!}$, $\theta_1,\theta_2,...,\theta_n$ are random numbers distributed as the mixing distribution $f(\theta)$. The remaining thing to do now is to simulate random numbers $\theta_1,\theta_2,...,\theta_n$ from the mixing distribution.

In Monte Carlo studies, there are various ways to simulate a random number from a given probability distribution. The method for simulation of the random numbers θ_i, i = 1, 2, ..., n from the mixing distribution $f(\theta)$ is chosen according to its suitability. When the cdf of the mixing distribution is available in closed form, such as in the case of the exponentiated Weibull and Weibull distributions, the inverse transform method uses the inverse of the cdf and is very fast. Otherwise, other methods such as the acceptance-rejection method or envelope rejection method can be used.

One of the drawbacks in the Monte Carlo simulation technique is in the computation time required. In practice, the number of points n which need to be simulated before the computation stabilizes may be unrealistically high. Therefore, variance reduction techniques are usually employed to accelerate convergence. To accelerate the convergence of estimator (4.3.3), we incorporate the use of one such technique, which is by substituting the usual (pseudo) random numbers used in the simulation to generate θ_i with a low discrepancy sequence of numbers instead. Since the low discrepancy sequence of numbers is a deterministic sequence, this method is also known as quasi-Monte Carlo method. Examples of low discrepancy sequence of numbers are the SOBOL sequence, Halton sequence and Niederreiter sequence (Sobol, 1998). In this study, we use the low F-discrepancy sequence of numbers (Fang & Wang, 1994) to reduce the sampling size required.

In the one-dimensional case,

$$u_i = \frac{2i-1}{2n}, \ i = 1, 2, 3, ..., n \qquad (4.3.4)$$

is the set of n points with the lowest F-discrepancy. Therefore, a Monte Carlo simulation algorithm to compute mixed Poisson probabilities is as follows.

i. Generate lowest-F discrepancy numbers (4.3.4).

ii. Compute $\theta_i = F^{-1}\left(\frac{2i-1}{2n}\right)$, where $F(\theta)$ is the cdf of the mixing distribution.

iii. Compute the estimator (4.3.3).

The algorithm can be modified accordingly for any arbitrary mixing distribution. For the Poisson-exponentiated Weibull probabilities, $F(\theta)$ in the algorithm is the cdf for the exponentiated Weibull distribution. The exponentiated Weibull's inverse cdf is given by $F^{-1}(\theta) = \beta[-\log(1-\theta^{\frac{1}{\eta}})]^{\frac{1}{\alpha}}$. In a similar manner, the probabilities for the Poisson-Weibull distribution can be computed using the inverse cdf of the Weibull distribution given by $F^{-1}(\theta) = \beta[-\log(1-\theta)]^{\frac{1}{\alpha}}$.

We exemplify the application of the Monte Carlo simulation technique in the evaluation of mixed Poisson probabilities by examining the computation of PIG, Poisson-lognormal and the newly proposed Poisson-exponentiated Weibull distribution. The results are presented in Tables 4.1, 4.2 and 4.3 respectively. Since the exact probabilities for the PIG distribution is known, we evaluate the accuracy (last column in Table 4.1) and measure of goodness-of-fit ε for this Monte Carlo approach. The goodness-of-fit measure is defined as $\varepsilon = \sum_k (p_k - p_k^*)^2 / p_k$, where p_k and p_k^* are the exact and approximate probabilities, respectively (Shaban, 1981). The first thirty probabilities are used to compute this measure and we obtain $\varepsilon = 6.3\times10^{-11}$ for the parameters in Table 4.1.

Table 4.1: PIG probabilities evaluated using (a) direct computation from formula (3.2.2), and (b) Monte Carlo estimator (4.3.3), for $\alpha = \theta = 0.5$

k	**Pr(*X* = *k*) Method (a)**	**Pr(*X* = *k*) Method (b)**	**Accuracy (%)**
2	0.02024465270	0.02024569286	99.99
3	0.00534233891	0.00534255198	99.99
4	0.00169584113	0.00169587264	99.99
5	0.00059771810	0.00059772521	99.99
10	0.00000640132	0.00000640135	99.99
15	0.00000010775	0.00000010775	99.99
20	0.0000000218	0.0000000218	99.99

Table 4.2: Poisson-lognormal probabilities evaluated using (a) numerical integration, and (b) Monte Carlo estimator (4.3.3), for μ = 2 and σ = 0.5

k	**Pr(*X* = *k*) Method (a)**	**Pr(*X* = *k*) Method (b)**
2	0.048659495	0.048659479
3	0.070620703	0.070620708
4	0.086194013	0.086193899
5	0.093741204	0.093741243
10	0.060678721	0.060678737
15	0.022880638	0.022880646
20	0.007610721	0.007610723

Table 4.3: **Poisson-exponentiated Weibull probabilities evaluated using (a) numerical integration, and (b) Monte Carlo estimator (4.3.3), for $\alpha = \beta$ = 0.5 and η = 2**

k	**Pr(*X* = *k*) Method (a)**	**Pr(*X* = *k*) Method (b)**
2	0.1094796399	0.1094796412
3	0.0653499921	0.0653499920
4	0.0418245007	0.0418245007
5	0.0281190033	0.0281190033
10	0.0059209016	0.0059209016
15	0.0018245689	0.0018245689
20	0.0006869179	0.0006869179

4.4 Parameter Estimation

Given the observations from the sample of interest, there are many parameter estimation methods for discrete distributions. Maximum likelihood (ML) estimation is a very popular approach due to its desirable properties such as efficiency, consistency and asymptotic normality. However, ML estimation has a major drawback as it may lead to intractable nonlinear likelihood equations. For mixed Poisson distributions, we can use global optimization routines such as the simulated annealing algorithm or an EM-type algorithm discussed by Karlis (2005). Although in general the EM-type algorithm can be adapted for all mixed Poisson distributions, we found that its performance is suboptimal for the generalized Sichel distribution. Therefore, the simulated annealing algorithm is used instead for parameter estimation in the generalized Sichel distribution.

4.4.1 Simulated Annealing

For the generalized Sichel distribution proposed in Chapter 3.3, the ML estimates of the unknown parameters $\boldsymbol{\omega} = (a, b, \delta, \lambda)$ of the generalized Sichel distribution is defined as $\hat{\boldsymbol{\omega}} = (\hat{a}, \hat{b}, \hat{\delta}, \hat{\lambda})^T = \arg\max_{\boldsymbol{\omega}} \log L(\boldsymbol{\omega})$, where $\log L$ is the log-likelihood function given by $\log L = \sum_{k=0}^{\infty} f_k \cdot \log[\Pr(X = k)]$ and f_k is the observed frequency of count k in the sample. In order to obtain the global maximum of the log-likelihood function analytically, we need to derive the partial derivatives of

$$\log L = \sum_{k=0}^{\infty} f_k \cdot \left\{ -\log\left(\frac{2}{\delta}\right) - \frac{\lambda}{2\delta}\log\left(\frac{b}{a}\right) - \log K_{\lambda/\delta}(2\sqrt{ab}) - \log k! + \log h(\theta) \right\}$$

where $h(\theta) = \int_0^{\infty} e^{-\theta}\theta^{k+\lambda-1}\exp(-a\theta^{\delta} - b\theta^{-\delta})d\theta$. We find that the likelihood equations are intractable, so we use the simulated annealing algorithm (Goffe, Ferrier & Rogers, 1994) to obtain the global maximum of the log-likelihood function.

4.4.2 Poisson-Weibull Distribution: ML Estimation via EM Algorithm

The pmf of the Poisson-Weibull distribution is not in closed form thus we are not able to derive the ML estimates of its parameters analytically. In the previous section, we use the simulated annealing algorithm to obtain the ML estimates of the generalized Sichel distribution. Although the simulated annealing algorithm is able to solve for a global maximum even for functions with many local maxima, it has a disadvantage of being time consuming to run. Since the Weibull distribution has a relatively simple pdf and ML estimator, the EM algorithm is a viable approach for obtaining ML estimates of the Poisson-Weibull parameters.

In Karlis' (2005) formulation for EM algorithm in the context for mixed Poisson distributions, the data sample of size *n* is treated as the complete data $Y_i = (X_i, \theta_i)$, *i* = 1, 2, ..., *n*, which consists of the observed data points X_i and the unobserved realizations θ_i of the parameter for each point X_i. As discussed in Section 4.1, θ is distributed as the mixing distribution $f(\theta; \boldsymbol{\tau})$ with parameter vector $\boldsymbol{\tau}$. At the E-step of the (*k*+1)-th iteration we use the estimates $\boldsymbol{\tau}^{(k)}$ from the *k*-th iteration to calculate pseudo-values $E(h_j(\theta) \mid X_i, \boldsymbol{\tau}^{(k)})$, for *i* = 1, 2, ..., *n*, *j* = 1, 2, ..., *m*, where $h_j(.)$ are certain functions depending on the mixing distribution. The expectation is taken with respect to the conditional distribution $f(Y; X, \boldsymbol{\tau}^{(k)})$. Then, at the M-step, using the pseudo-values from E-step, we maximize $Q(\boldsymbol{\tau}; \boldsymbol{\tau}^{(k)}) = E(\log p(Y \mid \boldsymbol{\tau}) \mid X, \boldsymbol{\tau}^{(k)})$ over $\boldsymbol{\tau}$.

The details of the EM algorithm for various mixed Poisson distributions such as the negative binomial, Poisson-inverse Gaussian and Poisson-lognormal distributions can be found in Karlis (2005). We follow the same approach for the Poisson-Weibull distribution and the algorithm is as follows:

E-step. Calculate the pseudo-values

$$t_i = E(\theta_i^{\alpha} \mid x_i) = \frac{\int e^{-\theta_i} \theta_i^{x_i+\alpha_{old}-1} (\theta_i^{\alpha}) \exp\left(-\left(\frac{\theta_i}{\beta_{old}} \right)^{\alpha_{old}} \right) d\theta}{\int e^{-\theta_i} \theta_i^{x_i+\alpha_{old}-1} \exp\left(-\left(\frac{\theta_i}{\beta_{old}} \right)^{\alpha_{old}} \right) d\theta} \qquad (4.4.2.1)$$

$$s_i = E(\log \theta_i \mid x_i) = \frac{\int e^{-\theta_i} \theta_i^{x_i+\alpha_{old}-1} (\log \theta_i) \exp\left(-\left(\frac{\theta_i}{\beta_{old}} \right)^{\alpha_{old}} \right) d\theta}{\int e^{-\theta_i} \theta_i^{x_i+\alpha_{old}-1} \exp\left(-\left(\frac{\theta_i}{\beta_{old}} \right)^{\alpha_{old}} \right) d\theta} \qquad (4.4.2.2)$$

Numerical integration is used to evaluate the integrals in the expectations.

M-step. From the likelihood function of the Weibull distribution, the ML estimate of β is given as $\hat{\beta} = \left(\frac{1}{n} \sum_{i=1}^{n} x_i^{\alpha} \right)^{\frac{1}{\alpha}}$ and the ML estimate of α satisfies $\frac{n}{\alpha} - \frac{n}{\sum_{i=1}^{n} x_i^{\alpha}} \sum_{i=1}^{n} x_i^{\alpha} \log(x_i) + \sum_{i=1}^{n} \log(x_i) = 0$. As such, we first compute $\beta_{new} = \left(\frac{1}{n} \sum_{i=1}^{n} t_i \right)^{\frac{1}{\alpha_{old}}}$ and subsequently we obtain α_{new} by numerically maximizing

$Q = n \log \alpha - n\alpha \log \beta_{new} + (\alpha - 1) \sum_{i=1}^{n} s_i - \left(\frac{1}{\beta_{new}} \right)^{\alpha} \sum_{i=`}^{n} t_i$ with respect to α.

The iterations are stopped if the relative change in the log-likelihood between two iterations is less than 10^{-10}. This is the same criteria adopted by Karlis (2005). The issue of choosing an appropriate termination criteria is common with EM algorithm and other termination criteria have been given by Seidel, Mosler and Alker (2000) and Karlis (2001).

Although the EM algorithm is a simple algorithm, it does not guarantee that the ML estimates found correspond to the global maximum of the likelihood function. This issue has been discussed by many authors, including Wu (1983) and Chung and Lindsay (2015). Two of the factors which cause the EM algorithm to fail to return the global maximum of the likelihood function is the termination criteria discussed in the preceding paragraph and choice of initial values. We adopt the standard practice of using multiple sets of randomly generated initial values to yield a higher confidence that the algorithm converges to the global maximum. Karlis and Xekalaki (2003) have conducted a study on the choice of the initial values and their affects on the ML estimation for finite mixtures.

4.5 Hypothesis Testing

Hypothesis testing is performed to determine the significance of the parameter δ in the generalized Sichel distribution and the parameter η in the Poisson-exponentiated Weibull distribution.

When η = 1, the Poisson-exponentiated Weibull distribution reduces to the Poisson-Weibull distribution. In order to test the significance of η, we perform the test of hypotheses

$$H_0 : \eta = 1 \quad \text{vs} \quad H_A : \eta \neq 1. \tag{4.5.1}$$

The difference between the Poisson-exponentiated Weibull and the Poisson-Weibull distributions is most prominent in the left tail probabilities. Therefore, we name the hypothesis test as the left-tail Weibull test. Since both the null and alternative hypotheses involve numerical integration, we use the likelihood ratio test for hypothesis testing. The likelihood ratio test statistic for (4.5.1) is

$$T_{LR} = 2[\log L(\hat{\alpha}, \hat{\beta}, \hat{\eta} \mid x_1, ..., x_n) - \log L(\hat{\alpha}_0, \hat{\beta}_0, 1 \mid x_1, ..., x_n)],$$

where $\log L(\hat{\alpha},\hat{\beta},\hat{\eta} \mid x_1,...,x_n)$ is the log-likelihood function of the Poisson-exponentiated Weibull distribution evaluated at the ML estimates and $\log L(\hat{\alpha}_0,\hat{\beta}_0,1 \mid x_1,...,x_n)$ is the log-likelihood function of the Poisson-exponentiated Weibull distribution evaluated at the restricted ML estimates. As η = 1 is an interior point in the parameter space, T_{LR} has an asymptotic chi-square distribution with one degree of freedom.

In the case of the generalized Sichel distribution, δ = 1 corresponds to the three-parameter Sichel distribution. As such, we name the hypothesis testing as the Sichel test. The null hypothesis H_0 and alternative hypothesis H_A can be written as

$$H_0 : \delta = 1 \quad \text{vs} \quad H_A : \delta \neq 1 \tag{4.5.2}$$

The distribution in the null hypothesis is the Sichel distribution which is relatively simple. In view of this, we perform the score test which requires only the restricted ML estimates corresponding to the Sichel distribution. Its test statistic for (4.5.2) is

$$T_{RS} = \mathbf{U}(\boldsymbol{\omega}_0)^T \mathbf{J}^{-1}(\boldsymbol{\omega}_0)\mathbf{U}(\boldsymbol{\omega}_0)$$

where the score vector $\mathbf{U}(\cdot) = (\partial \log L / \partial a, \partial \log L / \partial b, \partial \log L / \partial \delta, \partial \log L / \partial \lambda)^T$ is being evaluated at the null hypothesis, i.e. $\boldsymbol{\omega}_0 = (\hat{a}_0, \hat{b}_0, 1, \hat{\lambda}_0)^T$ are the restricted maximum likelihood estimates. The partial derivatives $\partial \log L / \partial a$, $\partial \log L / \partial b$, $\partial \log L / \partial \delta$ and $\partial \log L / \partial \lambda$ are given by (4.5.3), (4.5.4), (4.5.5) and (4.5.6), respectively. In general, $\mathbf{J}(\boldsymbol{\omega}_0)$ is either the expected or observed information matrix, also evaluated at the null hypothesis. In our case for the generalized Sichel distribution, we use the observed information matrix because the expected information matrix is intractable. The observed information matrix is the matrix of second partial derivatives of the log-likelihood function and its *r,s*-th element is obtained as $\{\mathbf{I}(\boldsymbol{\omega})\}_{rs} = \left[-\frac{\partial^2 \log L}{\partial \omega_r \partial \omega_s} \right]$, where ω_p is the *p*-th element in the parameter vector $\boldsymbol{\omega} = (a, b, \delta, \lambda)$. The score test statistic T_{RS} has an asymptotic chi-square distribution with one degree of freedom.

In order to obtain the likelihood score equations, we derive the partial derivatives of the log-likelihood function using the generalized Sichel pmf. As such,

$$\log L = \sum_{k=0}^{\infty} f_k \cdot \left\{ -\log\left(\frac{2}{\delta}\right) - \frac{\lambda}{2\delta} \log\left(\frac{b}{a}\right) - \log K_{\lambda/\delta}(2\sqrt{ab}) - \log k! + \log h(\theta) \right\}$$

where $h(\theta) = \int_0^{\infty} e^{-\theta} \theta^{k+\lambda-1} \exp(-a\theta^{\delta} - b\theta^{-\delta}) d\theta$.

Then the partial derivatives are

$$\frac{\partial \log L}{\partial a} = \sum_{k=0}^{\infty} f_k \left\{ \frac{\lambda}{2a\delta} - \frac{\partial}{\partial a} \log K_{\lambda/\delta}(2\sqrt{ab}) + \frac{\partial}{\partial a} \log h(\theta) \right\} \quad (4.5.3)$$

$$\frac{\partial \log L}{\partial b} = \sum_{k=0}^{\infty} f_k \left\{ -\frac{\lambda}{2b\delta} - \frac{\partial}{\partial b} \log K_{\lambda/\delta}(2\sqrt{ab}) + \frac{\partial}{\partial b} \log h(\theta) \right\} \quad (4.5.4)$$

$$\frac{\partial \log L}{\partial \delta} = \sum_{k=0}^{\infty} f_k \left\{ \frac{1}{\delta} + \frac{\lambda}{2\delta^2} \log\left(\frac{b}{a}\right) - \frac{\partial}{\partial \delta} \log K_{\lambda/\delta}(2\sqrt{ab}) + \frac{\partial}{\partial \delta} \log h(\theta) \right\} \quad (4.5.5)$$

$$\frac{\partial \log L}{\partial \lambda} = \sum_{k=0}^{\infty} f_k \left\{ -\frac{1}{2\delta} \log\left(\frac{b}{a}\right) - \frac{\partial}{\partial \lambda} \log K_{\lambda/\delta}(2\sqrt{ab}) + \frac{\partial}{\partial \lambda} \log h(\theta) \right\} \quad (4.5.6)$$

where

$$\frac{\partial}{\partial a} \{h(\theta)\} = \int_0^{\infty} (-\theta^{\delta}) e^{-\theta} \theta^{k+\lambda-1} \exp(-a\theta^{\delta} - b\theta^{-\delta}) d\theta,$$

$$\frac{\partial}{\partial b}\{h(\theta)\} = \int_0^\infty (-\theta^{-\delta})e^{-\theta}\theta^{k+\lambda-1}\exp(-a\theta^{\delta} - b\theta^{-\delta})d\theta\,,$$

$$\frac{\partial}{\partial \delta}\{h(\theta)\} = \int_0^\infty (\log\theta)(-a\theta^{\delta} + b\theta^{-\delta})e^{-\theta}\theta^{k+\lambda-1}\exp(-a\theta^{\delta} - b\theta^{-\delta})d\theta\,,$$

$$\frac{\partial}{\partial \lambda}\{h(\theta)\} = \int_0^\infty (\log\theta)e^{-\theta}\theta^{k+\lambda-1}\exp(-a\theta^{\delta} - b\theta^{-\delta})d\theta\,,$$

and the derivatives of the modified Bessel function of the third kind is obtained by differentiating $K_{\lambda/\delta}(2\sqrt{ab}) = \int_0^\infty \exp(-2\sqrt{ab}\cosh t)\cosh\left(\frac{\lambda}{\delta}t\right)dt$. A program written in MATLAB to compute the score test statistic is given in Appendix B.

4.6 Applications

To examine the suitability of the mixed Poisson distributions for zero-inflated, over dispersed and long-tailed data sets, we fit one simulated data set and four well-known data sets from the literature with some mixed Poisson distributions and the zero-inflated Poisson distribution. The four real data sets are from diverse fields of insurance, medicine and sales and marketing.

The ML estimates together with the maximized log-likelihoods and AIC values are presented in Table 4.4. The standard error for the best fitting model is provided. For the parameters of the generalized Sichel distribution, estimates of their standard error are obtained from the observed information matrix defined in Section 4.5. For each data set, based on the AIC values for model selection we select the best four models for goodness-of-fit test. The observed frequency and the fitted distributions are presented in Tables 4.5 - 4.9, together with the degrees of freedom, χ^2-statistic and *p*-values for the chi-square goodness-of-fit test. The χ^2-statistic and *p*-values reported in the tables are obtained after grouping to ensure that the expected frequency for each class is at least 5. The degree of freedom is equal to $(t-p-1)$ where t = number of classes and p = number of parameters.

4.6.1 Simulated data

In this section, we illustrate the application of the generalized Sichel distribution with a simulated data set with very long tail. The Malayan butterfly data is a well-known example in the literature on long-tailed data (Gupta & Ong, 2005). However, the frequencies after k = 25 are grouped hence the individual observations at the tail is lost. We simulate a long-tailed data using the estimated parameters of the Malayan butterfly data given in Gupta and Ong (2005). The mean and variance of the simulated data set are 10.6990 and 22.7307, respectively. A plot of the simulated data is given in Figure 4.1. The data set has a high zero count and a very long tail. The minimum value of the data set is 0, whilst the maximum is 224. During the model fitting, frequencies after k = 50 are grouped. From Table 4.4(a), the generalized Sichel, Poisson-exponentiated Weibull, Sichel and Poisson-lognormal distributions are selected for goodness-of-fit test. The model fitting results are presented in Table 4.5. For presentation purposes, observations after 20 have been grouped but the analysis is performed on the data used during the model fitting.

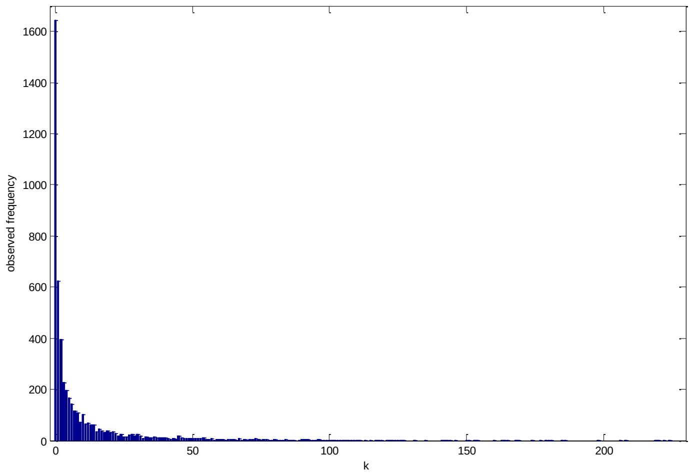


Figure 4.1: A plot of the frequency distribution of the simulated data

The generalized Sichel distribution and Poisson-exponentiated Weibull gives a satisfactory fit to the data in terms of chi-square goodness-of-fit statistic. The generalized Sichel distribution fits well on not only the observations at the right tail but also the zero and one counts.

Table 4.4: ML estimates, maximized log-likelihood and AIC

(a) Simulated data

Mixed Poisson Distribution	ML estimates (standard error)	Log-likelihood	AIC
Negative binomial	$\hat{\alpha}$ = 0.2989, $\hat{\beta}$ = 33.9736	-13697.53	27399.05
Poisson-lognormal	$\hat{\mu}$ = 0.6902, $\hat{\sigma}$ = 2.1803	-13687.63	27379.26
PIG	$\hat{\alpha}$ = 1.3198, $\hat{\theta}$ = 0.9983	-13836.28	27676.57
PGIG	$\hat{\alpha}$ = 0.3576, $\hat{\theta}$ = 0.9816, $\hat{\gamma}$ = -0.0774	-13654.18	27314.35
Poisson-Lindley	$\hat{\theta}$ = 0.2077	-17249.23	34502.47
Generalized Poisson-Lindley	$\hat{\alpha}$ = 0.0000, $\hat{\theta}$ = 0.1540	-14311.92	28627.84
Poisson-Weibull	$\hat{\alpha}$ = 0.5127, $\hat{\beta}$ = 4.7684	-14654.56	29313.12
Poisson-exponentiated Weibull	$\hat{\alpha}$ = 0.4156 (0.1403), $\hat{\beta}$ = 3.2881 (3.6174), $\hat{\eta}$ = 1.2340 (0.6601)	-13650.52	27305.04
Generalized Sichel	$\hat{a}$ = 0.2058, $\hat{b}$ = 0.0624, $\hat{\delta}$ = 0.5750, $\hat{\lambda}$ = 0.2632	-13650.23	27308.47

(b) Trobliger's data (1961) (Gathy & Lefèvre, 2010)

Mixed Poisson Distribution	ML estimates (standard error)	Log-likelihood	AIC
Negative binomial	$\hat{\alpha}$ = 1.1514, $\hat{\beta}$ = 0.1246	-10180.29	20364.57
Poisson-lognormal	$\hat{\mu}$ = -2.2651, $\hat{\sigma}$ = 0.8036	-10178.00	20360.00
PIG	$\hat{\alpha}$ = 1.2443, $\hat{\theta}$ = 0.2055	-10178.42	20360.83
PGIG	$\hat{\alpha}$ = 1.3138, $\hat{\theta}$ = 0.4060, $\hat{\gamma}$ = -1.8177	-10177.62	20361.25
Poisson-Lindley	$\hat{\theta}$ = 7.76669	-10181.02	20366.05
Generalized Poisson-Lindley	$\hat{\alpha}$ = 1.1375, $\hat{\theta}$ = 8.6533	-10180.34	20364.69
Poisson-Weibull	$\hat{\alpha}$ = 1.0720, $\hat{\beta}$ = 0.1473	-10179.64	20363.27
Generalized Sichel	$\hat{a}$ = 1.1407 (1.3912), $\hat{b}$ = 0.1514 (0.2447), $\hat{\delta}$ = 1.0560 (0.4198), $\hat{\lambda}$ = -1.9458 (0.8237)	-10177.60	20363.21
Zero-inflated Poisson	$\hat{\lambda}$ = 0.2538, $\hat{p}$ = 0.4348	-10190.58	20385.16

(c) Accident injuries data (Kadane et al., 2006)

Mixed Poisson Distribution	ML estimates (standard error)	Log-likelihood	AIC
Negative binomial	$\hat{\alpha}$ = 2.0361, $\hat{\beta}$ = 0.3474	-11485.48	22974.97
Poisson-lognormal	$\hat{\mu}$ = -0.5666, $\hat{\sigma}$ = 0.6620	-11463.00	22932.00
PIG	$\hat{\alpha}$ = 2.4604, $\hat{\theta}$ = 0.4330	-11466.82	22937.64
PGIG	$\hat{\alpha}$ = 2.7035, $\hat{\theta}$ = 0.9671, $\hat{\gamma}$ = -3.4594	-11454.22	22914.45
Poisson-Lindley	$\hat{\theta}$ = 1.8951	-11551.94	23105.88
Generalized Poisson-Lindley	$\hat{\alpha}$ = 1.9914, $\hat{\theta}$ = 3.15462	-11486.56	22977.12
Poisson-Weibull	$\hat{\alpha}$ = 1.4378, $\hat{\beta}$ = 0.7793	-11495.52	22995.04
Poisson-exponentiated Weibull	$\hat{\alpha}$ = 1.3963, $\hat{\beta}$ = 0.7748, $\hat{\eta}$ = 1.3340	-11489.76	22985.51
Generalized Sichel	$\hat{a}$ = 0.0027 (0.0067), $\hat{b}$ = 0.0715 (0.1170) $\hat{\delta}$ = 3.0000 (1.3142), $\hat{\lambda}$ = -2.1791 (0.2218)	-11450.78	22909.56
Zero-inflated Poisson	$\hat{\lambda}$ = 0.9135, $\hat{p}$ = 0.2257	-11613.88	23231.76

(d) Data on systemic adverse effects after vaccination (Rose et al., 2006)

Mixed Poisson Distribution	ML estimates (standard error)	Log-likelihood	AIC
Negative binomial	$\hat{\alpha}$ = 1.5268; $\hat{\beta}$ = 0.9870	-6740.60	13485.21
Poisson-lognormal	$\hat{\mu}$ = 0.1372; $\hat{\sigma}$ = 0.7469	-6762.87	13529.75
PIG	$\hat{\alpha}$ = 2.5208; $\hat{\theta}$ = 0.6782	-6760.27	13524.53
PGIG	$\hat{\alpha}$ = 0.0000, $\hat{\theta}$ = 0.4967, $\hat{\gamma}$ = 1.5268	-6740.60	13487.21
Poisson-Lindley	$\hat{\theta}$ = 0.9941	-6745.99	13493.99
Generalized Poisson-Lindley	$\hat{\alpha}$ = 1.2946, $\hat{\theta}$ = 1.1654	-6739.00	13482.01
Poisson-Weibull	$\hat{\alpha}$ = 1.2669; $\hat{\beta}$ = 1.6226	-6738.27	13480.53
Poisson-exponentiated Weibull	$\hat{\alpha}$ = 1.6210 (0.2271), $\hat{\beta}$ = 2.1305 (0.2415), $\hat{\eta}$ = 0.6428 (0.1391)	-6736.65	13479.29
Generalized Sichel	$\hat{a}$ = 0.1756, $\hat{b}$ = 0.0000 $\hat{\delta}$ = 1.8282, $\hat{\lambda}$ = 0.9879	-6736.70	13481.40

(e) Quarterly sales data (Shmueli et al., 2005)

Mixed Poisson Distribution	ML estimates (standard error)	Log-likelihood	AIC
Negative binomial	$\hat{\alpha}$ = 1.5817; $\hat{\beta}$ = 2.2505	-7526.61	15057.22
Poisson-lognormal	$\hat{\mu}$ = 0.9853; $\hat{\sigma}$ = 0.7709	-7562.27	15128.54
PIG	$\hat{\alpha}$ = 3.3987; $\hat{\theta}$ = 0.8394	-7562.22	15128.43

Table 4.4, continued

Mixed Poisson Distribution	ML estimates (standard error)	Log-likelihood	AIC
PGIG	$\hat{\alpha}$ = 0.0000, $\hat{\theta}$ = 0.6924, $\hat{\gamma}$ = 1.5817	-7526.61	15059.22
Poisson-Lindley	$\hat{\theta}$ = 0.4714	-7528.08	15058.16
Generalized Poisson-Lindley	$\hat{\alpha}$ = 1.1661, $\hat{\theta}$ = 0.5132	-7524.96	15053.93
Poisson-Weibull	$\hat{\alpha}$ = 1.2984 (0.0297); $\hat{\beta}$ = 3.8505 (0.0574)	-7525.40	15054.80
Poisson-exponentiated Weibull	$\hat{\alpha}$ = 1.2548, $\hat{\beta}$ = 3.6986, $\hat{\eta}$ = 1.0612	-7525.28	15054.56
Generalized Sichel	$\hat{a}$ = 0.2033, $\hat{b}$ = 0.0000 $\hat{\delta}$ = 1.2446, $\hat{\lambda}$ = 1.3334	-7525.34	15058.67

Table 4.4, continued

Table 4.5: **Fit of simulated data set**

k	**Observed frequency**	**Expected frequency**			
		Generalized Sichel	**Poisson-EW**	**Poisson-lognormal**	**Sichel**
0	1643	1643.24	1647.15	1559.80	1637.31
1	625	626.66	608.09	720.74	655.59
2	395	374.13	375.06	427.88	371.88
3	227	264.38	268.70	291.45	256.15
4	198	202.79	207.22	215.09	194.40
5	168	163.26	167.07	167.17	156.01
6	142	135.72	138.80	134.72	129.80
7	117	115.43	117.85	111.50	110.73
8	111	99.87	101.73	94.20	96.23
9	72	87.58	88.98	80.89	84.83
10	101	77.64	78.65	70.39	75.62
11	66	69.44	70.14	61.93	68.03
12	70	62.57	63.02	55.00	61.66
13	63	56.74	56.98	49.24	56.25
14	61	51.73	51.81	44.39	51.59
15	34	47.39	47.34	40.26	47.53
16	46	43.60	43.44	36.70	43.98
17	38	40.27	40.01	33.63	40.83
18	33	37.31	36.98	30.94	38.03
19	39	34.67	34.29	28.57	35.52
20	32	32.31	31.89	26.48	33.26
21 to 30	223	232.85	227.68	55.00	244.71
31 to 40	132	137.61	132.86	49.24	148.87
41 to 49	80	82.03	78.74	44.39	89.89
50 or more	284	280.77	285.54	40.26	271.30
Total	5000	5000	5000	5000	5000
Number of classes		51	51	51	51
Chi-square		54.4979	55.7434	137.3672	60.8280
d.f.		46	48	47	47
p-value		0.1827	0.2065	0.0000	0.0848

4.6.2 Real data

We present in Table 4.6 the fit for Tröbliger's data (1961) (as cited in Gathy & Lefèvre 2010) on the frequency of the number of claims. This data set has an 87% proportion of zeros. It has a mean of 0.1434 with standard deviation 0.4031, thus giving a dispersion index of 1.1328. From Table 4.4(b), with the exception of the Poisson-Lindley distribution, the AIC values for all mixed Poisson distributions considered are not significantly different. For further analysis, we fit the data to the generalized Sichel, PIG, Sichel and Poisson-Weibull distributions. The generalized Sichel and Sichel distributions provide a good fit amongst the four mixed Poisson distributions based on the *p*-value of the chi-square goodness-of-fit test. The Sichel distribution being a special case of the generalized Sichel distribution, fitting the generalized Sichel distribution eliminates the need for piece-wise treatment in the empirical modelling of the data.

The fit for data on number of injuries sustained in 10,000 accidents in the United States in 2001 (as cited in Kadane, Krishnan & Shmueli, 2006) is presented in Table 4.7. It has a mean of 0.7073, standard deviation 1.0020 thus yielding a dispersion index of 1.4194. Its proportion of zeros is at 54%. The generalized Sichel distribution gives a significantly better fit on this data in terms of its AIC values in Table 4.4(c) and chi-square goodness-of-fit statistic in Table 4.7, compared to the other mixed Poisson distributions considered here.

Table 4.8 gives the observed and expected frequencies by model and goodness-of-fit results for fitted models on the data on systemic adverse events for first four study injections in an anthrax vaccine absorbed (AVA) clinical trial study (Rose, Martin, Wannemuehler & Plikaytis, 2006). The data has sample mean 1.5070, standard deviation 1.7040 and dispersion index of 1.93, implying presence of over dispersion. We see that the Poisson-exponentiated Weibull distribution gives the best fit to the data.

Table 4.6: Fit of Trobliger's data (Gathy & Lefèvre, 2010)

Number of Claims	Observed Frequency	Expected Frequency			
		Generalized Sichel	Poison-Weibull	PIG	PGIG
0	20592	20593.33	20599.40	20597.30	20593.00
1	2651	2647.44	2624.52	2633.40	2647.52
2	287	291.39	314.01	303.63	291.69
3	41	38.52	36.42	38.37	38.50
4	7	6.52	4.14	5.34	6.50
5	0	1.35	0.46	0.80	1.35
6	1	0.32	0.05	0.13	0.32
>7	0	0.12	0.00	0.02	0.12
Total	23579	23579	23579	23579	23579
Number of classes		6	5	6	6
Chi-square		0.6165	5.5825	1.7271	0.6338
d.f.		1	2	3	2
p-value		0.4324	0.0613	0.6309	0.7284

Table 4.7: **Accident Injuries Data (Kadane et al., 2006)**

Number of accidents	Observed Frequency	Expected Frequency			
		Generalized Sichel	Poisson-lognormal	PIG	PGIG
0	5363	5389.30	5444.1919	5446.28	5408.54
1	3091	3025.28	2910.76	2900.77	2984.26
2	1008	1059.84	1085.87	1086.47	1072.46
3	348	332.94	367.25	372.34	345.44
4	105	111.96	123.35	126.44	114.36
5	46	43.16	42.76	43.60	41.47
6	19	18.89	15.54	15.35	16.81
7	9	9.04	5.96	5.52	7.59
8	7	4.56	2.41	2.02	3.76
9	2	2.37	1.02	0.75	2.02
10	1	1.25	0.46	0.28	1.15
> 11	1	1.39	0.43	0.18	2.13
Total	10000	10000	10000	10000	10000
Number of classes		9	9	8	10
Chi-square		6.0877	34.5882	40.1319	12.1989
d.f.		4	6	5	6
p-value		0.1927	0.0000	0.0000	0.0577

In Table 4.9, we present the observed and expected frequencies of the fitted models on the number of quarterly sales (Shmueli, Minka, Kadane, Borle & Boatwright, 2005). The mean is 3.5502 with standard deviation 3.3636, yielding dispersion index 3.19.

Both the Poisson-Weibull distribution and generalized Poisson-Lindley distributions give a satisfactory fit to the data as compared to the other mixed Poisson distributions considered.

All of the data sets cited here so far do not fit well to the zero-inflated Poisson distribution. This is due to a poor fit on the counts at the right tail of the data although it fits the zero counts very well. We also attempted to fit the data sets with the zero-inflated negative binomial (ZINB) distribution. However, the iterative method used to estimate the ZINB parameters failed to converge. This convergence failure is a common problem with the ZINB and it was also noted by Famoye and Singh (2006). Moreover, we observe that for all of the data sets cited here, the negative binomial predicted a higher frequency of zeros than which is observed, hence it may not be necessary to fit the ZINB model at all. As such, some mixed Poisson distributions such as the generalized Sichel distribution can serve as an alternative to model zero-inflated count data.

Table 4.8: Systemic adverse events after vaccination (Rose et al., 2006)

Number of adverse events	Observed Frequency	Expected Frequency			
		Generalized Sichel	Poisson-Weibull	Poisson-exponentiated Weibull	Generalized Poisson-Lindley
0	1437	1419.32	1414.65	1429.26	1418.11
1	1010	1024.63	1052.44	1026.71	1054.75
2	660	670.27	670.79	667.73	668.47
3	428	409.05	397.89	405.15	394.62
4	236	235.53	225.43	232.53	223.23
5	122	128.98	123.49	127.33	122.65
6	62	67.58	65.88	66.91	65.95
7	34	34.03	34.38	33.90	34.88
8	14	16.54	17.61	16.62	18.21
9	8	7.78	8.87	7.90	9.40
10	4	3.55	4.41	3.66	4.81
11	4	1.58	2.16	1.65	2.44
12	1	0.68	1.05	0.73	1.23
> 13	0	0.49	0.95	0.00	1.23
Total	4020	4020	4020	4020	4020
Number of classes		11	11	11	12
d.f.		6	8	7	9
Chi-square		3.8628	8.5174	4.2001	7.3941
p-value		0.6952	0.3846	0.7565	0.5962

Table 4.9: **Number of quarterly sales (Shmueli et al., 2006)**

Number of sales	Observed Frequency	Expected Frequency			
		Negative Binomial (Poisson-Gamma)	Generalized Poisson Lindley	Poisson-Weibull	Poisson-exponentiated Weibull
0	514	490.94	505.72	506.53	504.70
1	503	537.63	522.70	523.43	525.67
2	457	480.49	469.29	467.03	468.88
3	423	397.18	393.68	391.23	391.83
4	326	314.98	316.65	315.43	315.05
5	233	243.45	247.40	247.54	246.65
6	195	184.90	189.20	190.26	189.24
7	139	138.65	142.33	143.78	142.88
8	101	102.98	105.68	107.12	106.45
9	77	75.91	77.63	78.83	78.41
10	56	55.61	56.53	57.39	57.18
11	40	40.54	40.85	41.37	41.34
12	37	29.43	29.33	29.56	29.65
13	22	21.29	20.95	20.95	21.11
14	9	15.35	14.89	14.74	14.93
15	7	11.04	10.53	10.30	10.50
16	10	7.92	7.43	7.15	7.34
17	9	5.67	5.22	4.93	5.10
18	3	4.05	3.65	3.38	3.53
19	2	2.89	2.55	2.31	2.43
20	2	2.06	1.78	1.57	1.67
21	2	1.47	1.24	1.06	1.14
> 22	1	3.57	2.77	2.12	2.32
Total	3168	3168	3168	3168	3168
Number of classes		20	20	19	20
d.f.		17	17	16	16
Chi-square		17.3347	18.3217	15.6877	18.4206
p-value		0.4319	0.3688	0.4750	0.2998

Since the Poisson-Weibull and Poisson-exponentiated Weibull distributions give a similar fit to the quarterly sales data in Table 4.9, we perform the likelihood ratio test for the left-tail test (4.5.1) on the quarterly sales data. The test statistic obtained is 0.2357 with a *p*-value of 0.6273. At a significance level of $\alpha = 0.05$, we do not reject the null hypothesis and conclude that the Poisson-Weibull distribution gives a fit which is just as good as the Poisson-exponentiated Weibull distribution to this data.

In Tables 4.6 and 4.7, the *p*-value of the goodness-of-fit test implies that both the generalized Sichel and the Sichel distributions give a good fit to the data sets. For these two data sets, the hypothesis testing results for the Sichel test (4.5.2) are presented in

Table 4.10. At a significance level of $\alpha = 0.05$, the null hypothesis is not rejected for the Trobliger's data but is rejected for the accident injuries data. This conclusion corroborates the analysis of our model fitting results discussed earlier in this section.

Table 4.10: Score test results for Sichel test

Data Set	**Score Test Statistic**	***p*-value**
Tröbliger's data (1961) on number of claims	0.0566	0.8120
Accident injuries data	4.2257	0.0398

4.7 Conclusion

In the past, computational difficulties have hindered the use of many mixed Poisson distributions in empirical modelling. In this chapter, we apply a general computational approach for evaluating the mixed Poisson probabilities using a Monte Carlo simulation technique to compute Poisson-exponentiated Weibull probabilities. This method overcomes the computational hurdle which previously have hindered the application of many useful mixed Poisson distributions. An alternative is to use a numerical quadrature routine, of which we find the algorithm by Sermutlu and Eyyuboglu (2007) works well on the generalized Sichel probabilities.

Through the use of the Monte Carlo simulation technique together with an EM type algorithm for maximum likelihood estimation, we avoid the use of numerical methods and this allows many mixed Poisson distributions to be considered in model fitting. As we have shown in Section 4.6, mixed Poisson distributions such as generalized Sichel, Poisson-exponentiated Weibull and Poisson-lognormal are useful in modelling over dispersed data thus overcoming the computational issue is of significance.

Future work can be done to explore variance reduction techniques to speed up the Monte Carlo simulation technique in evaluating mixed Poisson probabilities. This would then enable the application of this technique for the generalized Sichel distribution, which at the moment takes a long time to compute with Monte Carlo simulation. The ML estimation problem for the generalized Sichel distribution and Poisson-Weibull distributions is treated differently due to the different structure of their respective mixing distributions. In principle, the EM algorithm can be applied on all mixed Poisson distributions but its implementation is less straightforward as the mixing distribution's likelihood function becomes more complicated. For instance, the M-step of the EM-type algorithm can be studied to incorporate conjugate gradient or quasi-Newton methods (Fletcher, 1987) so that this approach for ML estimation can be used for the class of mixed Poisson distributions where the ML equations of the mixing distribution is not in closed form.

CHAPTER 5: A FAMILY OF COUNT DISTRIBUTIONS ARISING FROM NON-EXPONENTIAL INTER-ARRIVAL TIMES

5.1 Introduction

A counting process is a stochastic point process $\{N(t), t \geq 0\}$ where $N(t)$ represents the total number of events that have occurred by time *t*. Let S_n denote the waiting time to (or arrival time of) the *n*-th event, and X_n denote the inter-arrival time or waiting time between the (*n* - 1)-st and the *n*-th event of this process. Therefore, $S_0 = 0$ and $S_n = \sum_{i=1}^{n} X_i$, $n \geq 1$. If the sequence of inter-arrival times $\{X_1, X_2, ...\}$ is iid as $f(x)$ with cdf $F(x)$, the counting process $\{N(t), t \geq 0\}$ is known as a renewal process. In a renewal process, the distribution function of S_n can be obtained as the *n*-fold convolution $F_n(x)$ of the distribution of X_i and $F_0(t) = 1$. From this point forward in this thesis, we assume that the inter-arrival times X_i is iid and we drop the index *i* from the notation thus *X* denotes the inter-arrival time.

In a renewal process, the number of events that has occurred up to a certain fixed time point *t* is closely related to the duration or inter-arrival times between these events. This relationship between the event counts $N(t)$ and inter-arrival times between events can be written as $N(t) \geq n \Leftrightarrow S_n \leq t$. As such, the pmf of the distribution of the event counts can be obtained as

$$\Pr\{N(t) = n\} = \Pr\{S_n \leq t\} - \Pr\{S_{n+1} \leq t\} = F_n(t) - F_{n+1}(t), \qquad (5.1.1)$$

where $F_n(x)$ is the cdf of S_n. Based on (5.1.1), the derivation of the count pmf is straightforward if $F_n(x)$ admits a closed form. However, very often $F_n(x)$ can only be expressed as convolution of integrals thus the count pmf becomes complicated.

In duration analysis, the hazard function of a distribution is used to capture duration dependence. A decreasing hazard function implies negative duration dependence. On the other hand, an increasing hazard function implies positive duration dependence, i.e. the waiting time is more likely to end the longer it lasts. The exponential distribution is a special distribution in the sense that it has a constant hazard function, known as the "memoryless" property.

The Laplace transform $\phi(s)$ of a function $f(x)$ is defined as $\phi(s) = \int_0^\infty e^{-sx} f(x)dx$, s a complex number. The Laplace transform exists for the function $f(x)$ defined over (0, ∞), whenever the integral converges. Since the inter-arrival times X_i's are iid, the Laplace transform of the waiting time $S_n = \sum_{i=1}^{n} X_i$ is simply the *n*-fold convolution of the Laplace transform of X_i. Consequently, the Laplace transform of the count distribution is derived as

$$\phi_n(s) = L(\Pr\{N(t) = n\}) = L(F_n(t) - F_{n+1}(t)) = \frac{1-\phi(s)}{\phi(s)}(\phi(s))^n \qquad (5.1.2)$$

where $\phi(s)$ is the Laplace transform of the inter-arrival time's pdf $f(x)$.

A trivial example of the relationship between the inter-arrival times and the event counts is when the inter-arrival times are exponentially distributed. Then the counting process is a Poisson process. Besides the exponential distribution, other positive continuous distributions especially lifetime distributions can be used to model the inter-arrival and waiting times in a renewal process. A survey on distributions that have been used to replace the exponential distribution is given in Section 5.2. In the subsequent sections, we derive the count distribution when the duration follows an inverse Gaussian distribution, convolution of two gamma distributions and a hyperexponential (finite mixture of exponentials, also known as Schuhl distribution) distribution. Winkelmann (1995) has shown that the dispersion of the count distribution in a stochastic process is related to the duration dependence of the inter-arrival distribution. If the hazard function is monotonic increasing, the count distribution will be under dispersed. A monotonic decreasing hazard function corresponds to an over dispersed count distribution. A constant hazard function, in the case of the exponential distribution, corresponds to an equidispersed Poisson distribution. Due to the flexibility of the hazard functions, the count distributions to be presented in Sections 5.3, 5.4 and 5.5 are over- and/or under dispersed. Some concluding remarks are given Section 5.6.

5.2 Literature Review

Lifetime distributions are a popular choice to model inter-arrival times in renewal processes. Banerjee and Bhattacharyya (1976) derived a new purchase count model by modelling the inter-purchase (inter-arrival of purchase incidence) times with an inverse Gaussian distribution. To account for population heterogeneity, they modelled the parameters of the inverse Gaussian distribution with a natural conjugate family. Compared to the negative binomial distribution, this compound inverse Gaussian count distribution is shown to give a better fit to a toothpaste purchase data set.

Winkelmann (1995) obtained a closed form expression for the count distribution with Erlangian inter-arrival times. In the same paper, the gamma distribution is used to replace the exponential distribution in modelling inter-arrival times. The Weibull distribution is a very popular model in reliability studies and McShane, Adrian, Bradlow and Fader (2008) derived the count distribution for Weibull duration in the regression context. In two separate occasions, Jose and Bindu have considered the Mittag-Leffler count distribution (Jose & Abraham, 2011) and Gumbel Type II (Jose & Abraham, 2013) count distribution. Recently Ong, Biswas, Peiris and Low (2015) considered a flexible generalized Weibull count distribution which is able to provide a better fit than the Weibull count distribution in modelling over dispersed and under dispersed count data. A summary of these count distributions is given in Table 5.1.

Table 5.1: Some existing count distributions in renewal theory

Inter-arrival time distribution	**Probability mass function (pmf) of corresponding count distribution**
Erlang	$\Pr\{N(t)=n\}=e^{-\beta t}\sum_{i=0}^{\alpha-1}\frac{(\beta t)^{\alpha n+i}}{(\alpha n+i)!}$, n = 0, 1, 2, ...
Gamma	$\Pr\{N(t)=n\}=G(\alpha n,\beta t)-G(\alpha n+\alpha,\beta t)$, where $G(\alpha n,\beta t)=\frac{1}{\Gamma(n\alpha)}\int_0^{\beta t}u^{n\alpha-1}e^{-u}du$
Weibull	$\Pr\{N(t)=n\}=\sum_{j=n}^{\infty}\frac{(-1)^{j+n}(\lambda t^c)^j\alpha_j^n}{\Gamma(cj+1)}$, where $\alpha_j^0=\frac{\Gamma(cj+1)}{\Gamma(j+1)}$, j = 0, 1, 2, ... and $\alpha_j^{n+1}=\sum_{m=n}^{j-1}\alpha_m^n\frac{\Gamma(cj-cm+1)}{\Gamma(j-m+1)}$ for n = 0, 1, 2, ... for j = n+1, n+2, n+3, ...
Mittag-Leffler	$\Pr\{N(t)=n\}=\sum_{j=n}^{\infty}\frac{\binom{j}{n}(-1)^{(j-n)}t^{j\alpha}}{\Gamma(1+j\alpha)}$
Gumble Type II	$\Pr\{N(t)=n\}=\sum_{j=n}^{\infty}\frac{(-1)^{(j+n)}(bt^{-a})^j\delta_j^n}{\Gamma(-aj+1)},a<0$, where $\delta_j^0=\frac{\Gamma(-aj+1)}{\Gamma(j+1)}$, j = 0, 1, 2, ... and $\delta_j^{n+1}=\sum_{m=n}^{j-1}\delta_m^n\frac{\Gamma(-aj+am+1)}{\Gamma(j-m+1)}$ for n = 0, 1, 2, ... for j = n+1, n+2, n+3, ...
Generalized Weibull	$\Pr\{N(t)=n\}=(a\alpha)^n\sum_{p=0}^{\infty}\frac{(-a/\lambda)^p}{\Gamma(\alpha(p+n)+1)}t^{\alpha(p+n)}c_n(p)$, where $c_n(p)=\sum_{q=0}^{p}\binom{\lambda-1}{q}\Gamma(\alpha(q+1))c_{n=1}(p-q),n\geq 1$ and $c_0(p)=\binom{\lambda}{p}\Gamma(\alpha p+1)$.

5.3 Inverse Gaussian Count Distribution

In this section we derive the count distribution when the inter-arrival times has an inverse Gaussian (IG) distribution (Johnson, Kotz & Balakrishnan, 1994). Our duration dependence-based approach is different from the work by Banerjee and Bhattacharyya (1976) in that their model accounts for heterogeneity in the population by assuming a compound inverse Gaussian distribution. The interpretation of the IG distribution as a first passage time distribution of Brownian motion with positive drift has resulted in the distribution being used as a lifetime distribution and duration model in various fields. The IG distribution has been used to model duration of a strike (Lancaster, 1972), length of hospital stays (Eaton & Whitmore, 1977) and employee service times (Whitmore, 1979). The pdf of the IG distribution is given by

$$f(x;\mu,\lambda)=\sqrt{\frac{\lambda}{2\pi}}x^{-\frac{3}{2}}\exp\left\{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right\} \tag{5.3.1}$$

for $x>0$, where μ, $\lambda > 0$. The hazard function (Chhikara & Folks, 1977) is

$$h(x)=\frac{(\lambda/2\pi x^3)^{1/2}\exp(-\lambda(x-\mu)^2/2\mu^2 x)}{\Phi(\sqrt{\lambda/x}(1-x/\mu))-e^{2\lambda/\mu}\Phi(-\sqrt{\lambda/x}(1+x/\mu))},$$

where Φ denotes the cdf of the standard normal distribution. There are several parameterizations of the IG distribution, but we use (5.3.1) because the parameters has meaningful interpretations $\mu=E(X)$ and λ is the scale parameter. It is a unimodal and positively skewed distribution. The shape of the distribution is determined by the ratio $\frac{\lambda}{\mu}$ and the pdf is highly skewed for moderate values of this ratio. Chhikara and Folks (1977) describes the failure rate of the IG distribution as an increasing function until it reaches a maximum point and subsequently the failure rate function decreases towards an asymptotic value of $\lambda/2\mu^2$.

When $\mu \to \infty$, we obtain a one-parameter limiting form of IG, known as the distribution of first passage time of drift-free Brownian motion. Its pdf is given as

$$f(x;\lambda) = \sqrt{\frac{\lambda}{2\pi}}\, x^{-\frac{3}{2}} \exp\left(-\frac{\lambda}{2x}\right)$$

with $x > 0$, where $\lambda > 0$ (Johnson et al., 1994). The expected value and variance of this distribution are infinite. On the other hand, when $\mu = 1$, the distribution is also known as the Wald distribution.

The Laplace transform of the inverse Gaussian distribution (Seshadri, 1983) is given as

$$\phi(s) = \exp\left\{\frac{\lambda}{\mu}\left(1 - \sqrt{1 + \frac{2s\mu^2}{\lambda}}\right)\right\}. \qquad (5.3.2)$$

Proposition 5.1 If the inter-arrival time (duration) has an inverse Gaussian distribution with pdf (5.3.1), the inverse Gaussian count distribution has pmf given by

$$\Pr\{N(t) = n\} = \sum_{k=0}^{\infty}\sum_{l=0}^{k} \frac{n^{k-l}}{(l+1)!(k-l)!}\left(\frac{\lambda}{\mu}\right)^{k+1} c_k(m), \qquad (5.3.3)$$

where $c_k(m) = \sum_{m=0}^{k+1}\binom{k+1}{m}(-1)^m \left(\sum_{\nu=0}^{\infty}\binom{\frac{m}{2}}{\nu}\frac{1}{\Gamma(1-\nu)}\left(\frac{2\mu^2}{\lambda t}\right)^{\nu}\right)$.

The derivation of expression (5.3.3) is given in Appendix E.

Since the IG duration model has an almost increasing failure rate, the IG count distribution is more suited to model under dispersed data, although it may also be able to model over dispersion. The pmf plots of the inverse Gaussian count distribution and the Poisson distribution is given in Figure 5.1. For comparison purposes, the mean for both distributions is set to 2 ($E(N) = 2$).

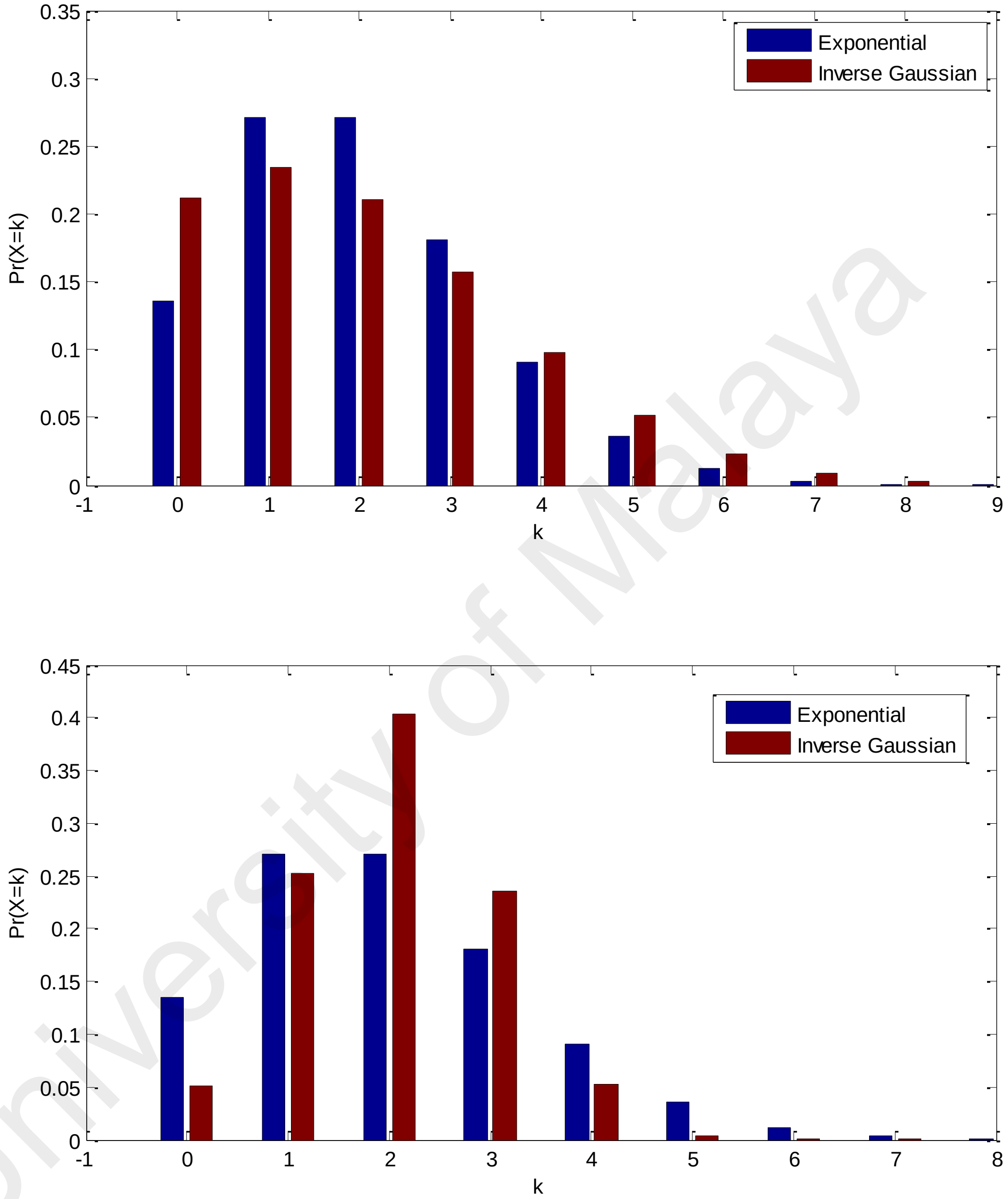


Figure 5.1: Probability functions for the Poisson and inverse Gaussian count distribution; (top) λ = 0.17, μ = 1 (over dispersion), (bottom) λ = 1, μ = 0.438 (under dispersion)

5.4 Count Distribution for Convolution of Two Gamma Duration

The gamma duration model has been used by Winkelmann (1995) to derive a gamma count distribution for regression modelling. The gamma density function is defined as $f(x) = \frac{\beta}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \alpha, \beta > 0$. We find that the gamma count distribution is a satisfactory model for many over dispersed and under dispersed data. It is then of interest to examine the convolution of two gamma distributions as a duration model. This generalization is motivated by an interpretation that the inter-arrival time in a duration model may be determined by two variables both of which are gamma distributed. In discussing the use of convolution distributions as duration models, Ickowicz and Sparks (2015) provide an example whereby the length of stay in a hospital could possibly be determined by patient's need and the hospital's discharging process.

If we represent the inter-arrival time X as a sum of two independent gamma random variables, then X has a convolution of two gamma distributions. A brief review on the convolution of two gamma density functions can be found in Johnson et al. (1994). We shall adapt the density function given in Moschopoulos (1985) for the sum of n independent gamma random variables, which is derived from the n-convolutions of the moment generating function. Let $X = X_1 + X_2$, where X_i, i = 1, 2, are distributed as gamma with parameters α_i and β_i respectively. We obtain the density function of X as

$$f(x;\rho,\beta_1) = \left(\frac{\beta_1}{\beta_2}\right)^{\alpha_2} \sum_{k=0}^{\infty} \frac{\delta_k x^{\rho+k-1} \exp\left(-\frac{y}{\beta_1}\right)}{\Gamma(\rho+k)\beta_1^{\rho+k}}, \qquad (5.4.1)$$

for $x > 0$, where $\beta_1 = \min(\beta_1, \beta_2)$, $\rho = \alpha_1 + \alpha_2$, $\delta_{k+1} = \frac{1}{k+1}\sum_{i=1}^{k+1} i\gamma_i \delta_{k+1-i}$ with $\delta_0 = 1$ for k = 0, 1, 2, ..., and $\gamma_k = \left\{\alpha_2\left(1-\frac{\beta_1}{\beta_2}\right)^k\right\}$.

Proposition 5.2 If the inter-arrival time (duration) has a convolution of two gamma distributions with pdf (5.4.1), the count distribution has pmf given by

$$\Pr\{N(t)=n\} = C_n(t,\alpha_1,\alpha_2,\beta_1,\beta_2) - C_{n+1}(t,\alpha_1,\alpha_2,\beta_1,\beta_2) \qquad (5.4.2)$$

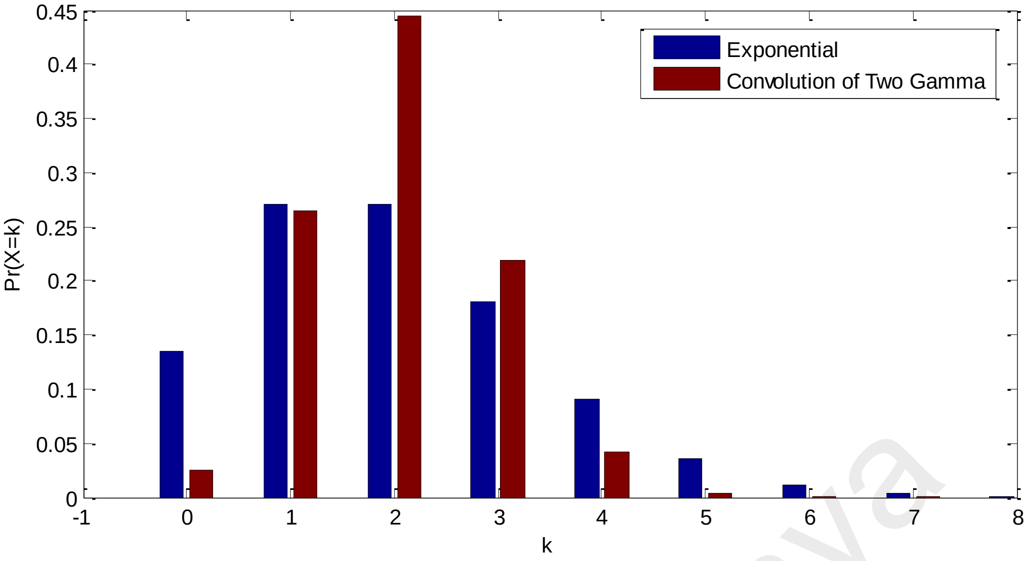
where

$$C_n(t,\alpha_1,\alpha_2,\beta_1,\beta_2) = (\beta_1^{\alpha_1}\beta_2^{\alpha_2})^n \left\{ \frac{t^{n(\alpha_1+\alpha_{20})}}{\Gamma(1+n(\alpha_1+\alpha_2)} \Phi_2(n\alpha_1, n\alpha_2; 1+n(\alpha_1+\alpha_2); -\beta_1 t, -\beta_2 t) \right\}$$

and $\Phi_2(b,b';c;w,z) = \sum_{k,l=0}^{\infty} \frac{(b)_k (b')_l}{(c)_{k+l}} \frac{w^k z^l}{k!l!}$.

The derivation of expression (5.4.2) is given in Appendix F.

The convolution of two gamma distributions has an increasing hazard function when its two component distributions has an increasing hazard function, i.e. both α_1 and α_2 are greater than 1. In other cases, convolution of two gamma distributions may give rise to a distribution with increasing or decreasing hazard function. Therefore, the convolution of two gamma count model can also model both over dispersion and under dispersion relative to the Poisson distribution, with some added flexibility than the simple gamma count model. Figure 5.2 compares the probability functions of the convolution of two gamma count distribution with a Poisson distribution.



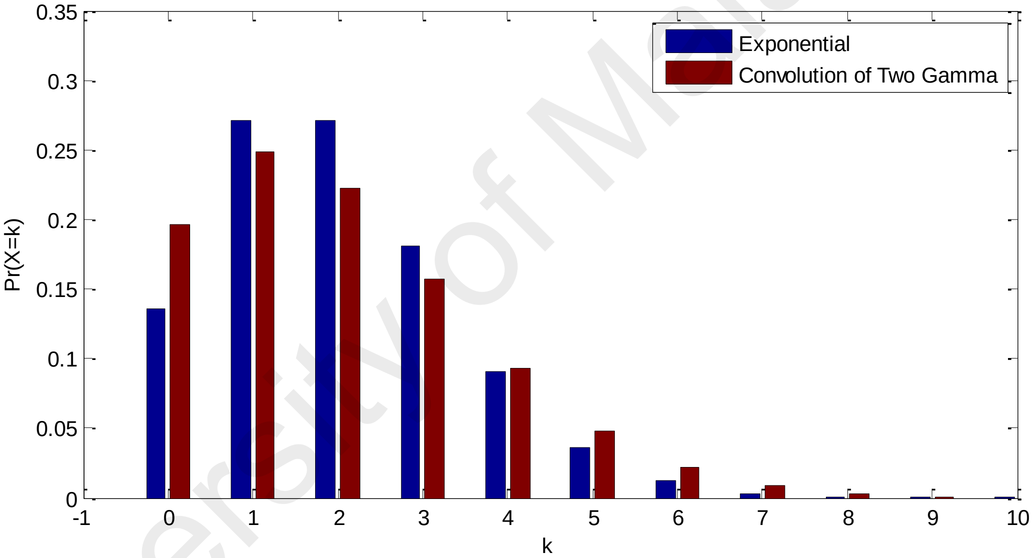


Figure 5.2: Probability functions for the Poisson and convolution of two gamma count distribution; (top) α_1 = 1.5, α_2 = 1.9 (under dispersion), (bottom) α_1 = 0.2, α_2 = 0.5 (over dispersion)

The convolution of two gamma distributions nests the special case of convolution of two exponential distributions, i.e. when $\alpha_1 = \alpha_2 =$ 1. This two-component hypoexponential count distribution with parameters β_1 and β_2 can model under dispersion and Figure 5.3 compares its probability function with a Poisson distribution.

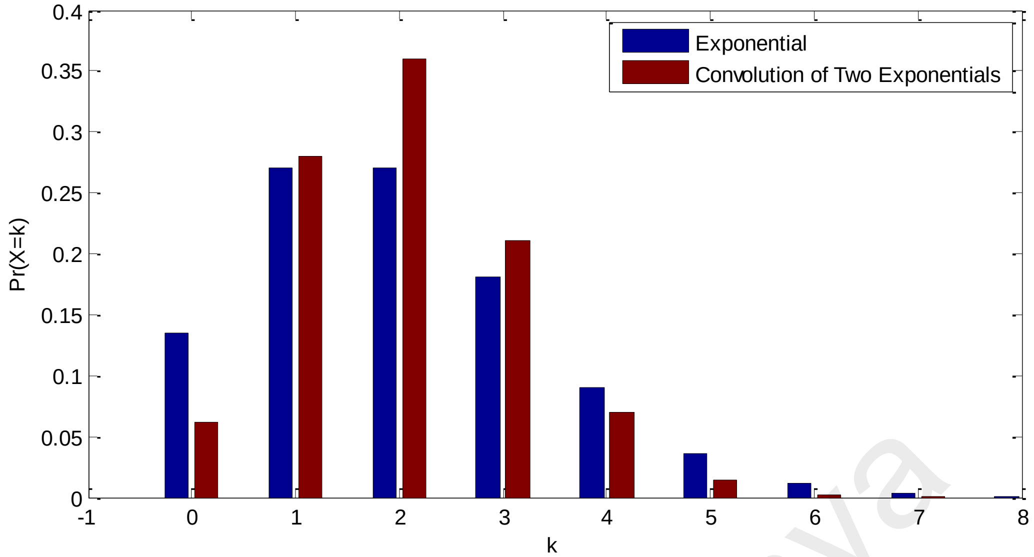


Figure 5.3: Probability function for the Poisson and convolution of two exponentials count distribution; $\beta_1 = 4.2$, $\beta_2 = 4.85$ (under dispersion)

5.5 Count Distributions with Finite Mixture Inter-arrival Times

In this section we derive the count distribution when the duration has a finite mixture of exponential model. The concept of finite mixture is used to model a population which consists of subpopulations. As such, using finite mixture distributions as an underlying duration model in a stochastic process can account for the heterogeneity present in the population when modelling the event counts. An immediate example is the possible existence of two subpopulations of subjects in the study of health care utilization, i.e. those who rarely consults a specialist or those who sees a specialist regularly.

The hyperexponential distribution has been used to approximate the distribution of long-tailed inter-arrival times (Feldmann & Whitt, 1998), in comparison with the Weibull and Pareto distributions. Nair and Abdul (2010) derived several important properties of the two-component hyperexponential distribution in the context of renewal theory. It is a three-parameter distribution, which is relatively simple as compared to other finite mixture distributions.

The pdf of a finite mixture of exponentials is given by

$$f(x) = \sum_{i=1}^{k} p_i \lambda_i e^{-\lambda_i x} \tag{5.5.1}$$

where p_i is known as the phase probability, $0 < p_i < 1$, $\sum_{i=1}^{\infty} p_i = 1$ and $\lambda_i > 0$, $i = 1, 2, ..., k$. Specifically, we consider the case of $k = 2$, i.e. mixture of two exponential distributions which considerably simplifies (5.5.1) to

$$f(x) = p\lambda_1 e^{-\lambda_1 x} + (1-p)\lambda_2 e^{-\lambda_2 x}, \tag{5.5.2}$$

where $0 < p < 1$ and $\lambda_1, \lambda_2 > 0$. The hazard function is

$$h(x) = \frac{p\lambda_1 e^{-\lambda_1 x} + (1-p)\lambda_2 e^{-\lambda_2 x}}{pe^{-\lambda_1 x} + (1-p)e^{-\lambda_2 x}}. \tag{5.5.3}$$

We obtain its Laplace transform as

$$\phi(s) = \frac{p\lambda_1}{\lambda_1 + s} + \frac{(1-p)\lambda_2}{\lambda_2 + s}.$$

Nair and Abdul (2010) has derived the pdf and cdf of the waiting time S_n as

$$g(t) = \sum_{k=0}^{\infty} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \lambda_1{}^k \lambda_2{}^{n-k} t^{n-1} \phi_2(k, n-k, n-\lambda_1 t, -\lambda_2 t) \tag{5.5.4}$$

and

$$G_n(t) = \frac{1}{n} \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \lambda_1{}^k \lambda_2{}^{n-k} t^n \phi_2(k, n-k, n+1-\lambda_1 t, -\lambda_2 t) \tag{5.5.5}$$

where $\phi_2(.)$ is the hypergeometric series $\phi_2(\beta,\beta',\gamma,x,y)=\sum_{u=0}^{\infty}\sum_{v=o}^{\infty}\frac{(\beta)_u(\beta')_v x^u y^v}{(\gamma)_{u+v}u!v!}$.

Consequently, if the inter-arrival time (duration) has a hyperexponential distribution with pdf (5.5.2), the hyperexponential count distribution has pmf given by

$$P\{N(t)=n\}=G_n(t)-G_{n+1}(t),\ n=0,1,2,\ldots, \tag{5.5.6}$$

where $G_n(t)=\frac{1}{n}\sum_{k=0}^{n}\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}\lambda_1^{\,k}\lambda_2^{\,n-k}t^n\phi_2(k,n-k,n+1-\lambda_1 t,-\lambda_2 t)$ is as defined in (5.5.5).

The hyperexponential distribution has a monotone decreasing hazard function thus the hyperexponential count distribution (5.5.6) can model over dispersed data.

5.6 Conclusion

In this chapter, a family of count distributions is derived by considering some non-exponential inter-arrival times. The choice of the inverse Gaussian distribution, convolution of two gamma distributions and the hyperexponential distribution are motivated by their applications in modelling life times, the behaviour of their hazard function and their physical interpretation in modelling duration. By taking into consideration the hazard function of the inter-arrival times, this family of count distributions has the flexibility of being able to model under-, equi- and over dispersed count data.

CHAPTER 6: COMPUTATION OF PROBABILITIES AND STATISTICAL INFERENCE FOR NON-EXPONENTIAL DURATION COUNT DISTRIBUTIONS

6.1 Introduction

Count distributions arising from non-exponential inter-arrival times in a stochastic process are able to model under- and/or over dispersed data, depending on the hazard function of the inter-arrival times' distribution. However, with the exception of the Erlangian count distribution proposed by Winkelmann (1995), the pmf of these count distributions are complicated, often involves special mathematical functions or computationally intractable expression. For example, the generalized Weibull count probabilities derived by Ong, Biswas, Peiris and Low (2015) involves an infinite series and a gamma function $\Gamma(x)$ which tends to numerically overflow quickly.

In this chapter, a numerical inverse Laplace transform technique is proposed for computing the probabilities of count distributions arising from renewal processes with non-exponential duration. Using this method, computational restrictions which limits the use of many lifetime distributions to model inter-arrival times can be overcome.

The organization of this chapter is as follows. A literature review on the existing methods used in computing the count probabilities are given in Section 6.2. The implementation of the algorithm for the proposed technique, as well as a comparison on its accuracy with existing methods in the literature is discussed in Section 6.3. The proposed method is used for calculating the renewal function in Section 6.4. In Section 6.5, we use the proposed technique for fitting under- and over dispersed data sets with the family of count distributions proposed in Chapter 5. A hypothesis test is performed for the hyperexponential count distribution in Section 6.6. Finally, concluding remarks is given in Section 6.7.

6.2 Literature Review

McShane, Adrian, Bradlow and Fader (2008) and Jose and Abraham (2013) have used the polynomial expansion method to derive the count distribution for Weibull and Gumbel inter-arrival times, respectively. A different approach by From (2004) used a family of generalized Poisson distributions to approximate the renewal counting processes with Weibull, truncated normal and exponentiated Weibull inter-arrival times. Chaudhry, Yang and Ong (2013) used the method of roots and a Padè approximation method for computing the count probabilities for several inter-arrival times distributions.

An important concept in renewal processes is the renewal function or expected number of renewals $E\{N(t)\}$ which is defined as

$$H(t) = E[N(t)] = \sum_{i=1}^{\infty} F_i(t). \tag{6.2.1}$$

The Laplace transform of (6.2.1) is

$$L(E[N(t)]) = \frac{\phi(s)}{s}\frac{1}{(1-\phi(s))}, \ |\phi(s)| < 1.$$

There are many studies on the approximation of the renewal function. Using a generalized cubic splining algorithm which provides piecewise polynomial approximations to recursively-defined convolution integrals, Baxter, Scheuer, Blischke and McConalogue (1981) have tabulated the renewal function and variance function for renewal processes with gamma, inverse Gaussian, lognormal, truncated normal and Weibull inter-arrival times. However, they noted that the convergence of the algorithm is slow for some of the parameter values. Chaudhry et al. (2013) took a slightly different approach by using the probability function obtained from numerically inverting the

Laplace transform in rational function form to calculate the renewal function and variance of several count distributions. They obtained the distribution function, mean and variance of $N(t)$ using the method of roots for numerically inverting the Laplace transform when it can be expressed as a rational function. They also studied the Padè approximation method to obtain an approximate rational function for the Laplace transform when it is not a rational function. In addition, they used Padè approximation method prior to the method of roots when the Laplace transform could not be expressed as a rational function, such as in the case for gamma and inverse Gaussian distributions.

6.3 Computation of the Probabilities of Count Distribution

In this section, we discuss the proposed numerical inverse Laplace transform technique for computing the probabilities of count distribution. The relationship between the Laplace transform of the inter-arrival times' distribution and the count distribution has been discussed in Chapter 5. For easier reference, we reproduce equation (5.1.2) here as

$$\phi_n(s) = L(\Pr\{N(t) = n\}) = L(F_n(t) - F_{n+1}(t)) = \frac{1-\phi(s)}{\phi(s)}(\phi(s))^n, \qquad (6.3.1)$$

where $\phi(s)$ is the Laplace transform of the inter-arrival time's pdf $f(x)$.

The probability function of the counts can be recovered by numerically inverting the Laplace transform (6.3.1). Using this method, given the inter-arrival time distribution and its Laplace transform, we will be able to compute the corresponding count probabilities.

For some common functions, the inverse Laplace transforms $f(x)$ are readily available from existing tables (Erdélyi, 1953). Otherwise, there are various explicit formulae for inverting a Laplace transform $\phi(s)$, such as the Bromwich inversion integral formula and the Post-Widder inversion formula. In most cases, it is difficult to find an analytical expression for the inverse Laplace transform using these formula thus a numerical inversion is necessary. Abate and Valkó (2004) and Valkó (2015) have given a comprehensive review on the numerous methods for numerical inversion of Laplace transforms. In our study, we use a numerical inversion algorithm which is based on the Bromwich inversion integral and gives good results for smooth functions. The algorithm was originally proposed by Dubner and Abate (1968), improved by Abate and Whitt (1992) and was proposed by Abate and Whitt (1995) for numerically inverting Laplace transforms of probability distributions.

The basis of the proposed algorithm is the Bromwich inversion integral formula which is given as

$$f(x) = L^{-1}(\phi(s)) = \lim_{R\to\infty} \frac{1}{2\pi i} \int_{a-iR}^{a+iR} \phi(s) e^{sx} ds \,, \tag{6.3.2}$$

where a is another real number such that $a > s_0$, s_0 is the real part of the singularities of $\phi(s)$ and $i = \sqrt{-1}$. The algorithm is developed by first applying the trapezoidal rule to the integral in (6.3.2), and subsequently using a Fourier-series method for approximation. Based on the algorithm, we obtain the following formula for computing the count probabilities

$$\Pr\{N(t) = n\} = \frac{e^{A/2}}{2t} \operatorname{Re}\left(\phi_n\left(\frac{A}{2t}\right)\right) + \frac{e^{A/2}}{t} \sum_{k=1}^{\infty} (-1)^k \operatorname{Re}\left(\phi_n\left(\frac{A + 2k\pi i}{2t}\right)\right), \tag{6.3.3}$$

where $\phi_n(.)$ is as defined in (6.3.1).

The convergence of the infinite sum in (6.3.3) can be accelerated by applying the well-known Euler's algorithm for alternating series. Therefore, the count probabilities (6.3.3) are approximated using the following formula

$$\Pr\{N(t)=n\} \approx \sum_{k=0}^{m} \binom{m}{k} 2^{-m} s_{p+k}(t), \tag{6.3.4}$$

where $s_p(t)$ is the p-th partial sum

$$s_p(t) = \frac{e^{A/2}}{2t} \operatorname{Re}\left(\phi_n\left(\frac{A}{2t}\right)\right) + \frac{e^{A/2}}{t} \sum_{k=1}^{p} (-1)^k \operatorname{Re}\left(\phi_n\left(\frac{A+2k\pi i}{2t}\right)\right).$$

The choice of the constant A affects the discretization error caused by using the trapezoidal rule. We use Abate and Whitt's (1995) suggestion to set A = 18.4, p = 38 and m = 11. The value of p may be increased when necessary. The algorithm can be implemented in programming languages which provides for complex number computation, such as MATLAB©.

6.3.1 Implementation

In order to implement the numerical Laplace transform inversion method for computing the count probabilities, knowledge of the inter-arrival time distribution's Laplace transform is necessary. The Laplace transform of the exponential, Erlang, gamma distributions, as well as of those proposed in Chapter 5, is listed in Table 6.1. The Laplace transform of the Weibull and generalized Weibull distributions cannot be in expressed in a closed form.

To illustrate the accuracy of this numerical Laplace transform inversion method, we apply it in calculating the count probabilities for generalized Weibull duration and Erlangian duration. The pmf of the Erlangian count distribution is given as (Winkelmann, 1995)

$$\Pr\{N(t)=n\}=e^{-\beta t}\sum_{i=0}^{\alpha-1}\frac{(\beta t)^{\alpha n+i}}{(\alpha n+i)!}$$

for n = 0, 1, 2, …, where α integer and β > 0. The pmf of the Erlangian count distribution is in closed form and computationally simple, hence there is no actual need to use the proposed method but it serves as a good example for comparison purpose.

Table 6.1: Laplace transforms

Distribution of the inter-arrival times	**Laplace Transform**
Exponential	$\phi(s)=\frac{\lambda}{\lambda+s}$, $\lambda>0$
Erlang	$\phi(s)=\left(\frac{\beta}{\beta+s}\right)^{\alpha}$, α integer and $\beta>0$
Gamma	$\phi(s)=\left(\frac{\beta}{\beta+s}\right)^{\alpha}$, $\alpha>0$ and $\beta>0$
Inverse Gaussian (Seshadri, 1983)	$\phi(s)=\exp\left\{\frac{\lambda}{\mu}\left(1-\sqrt{1+\frac{2s\mu^2}{\lambda}}\right)\right\}$, $\mu>0$ and $\lambda>0$
Convolution of two gamma distributions	$\phi(s)=\left(\frac{\beta_1}{\beta_1+s}\right)^{\alpha_1}\left(\frac{\beta_2}{\beta_2+s}\right)^{\alpha_2}$, $\alpha_i>0$ and $\beta_i>0$, $i=1,2$
Hyperexponential	$\phi(s)=\frac{p\lambda_1}{\lambda_1+s}+\frac{(1-p)\lambda_2}{\lambda_2+s}$, $0<p<1$, $\lambda_i>0$, $i=1,2$

The pmf of the generalized Weibull count distribution is defined as (Ong et al., 2015)

$$\Pr\{N(t)=n\}=(a\alpha)^n\sum_{p=0}^{\infty}\frac{(-a/\lambda)^p}{\Gamma(\alpha(p+n)+1)}t^{\alpha(p+n)}c_n(p) \qquad (6.3.1.1)$$

where $c_n(p)=\sum_{q=0}^{p}\binom{\lambda-1}{q}\Gamma(\alpha(q+1))c_{n-1}(p-q), n\geq 1$ and $c_0(p)=\binom{\lambda}{p}\Gamma(\alpha p+1)$, and a, α > 0, $x>0$ if $\lambda\leq 0$ and $0<x<(\lambda/a)^{1/\alpha}$ if $\lambda>0$. The Laplace transform of the generalized Weibull density function is approximated using Gauss-Laguerre quadrature since it is not available in closed form.

The comparison table is presented in Table 6.2. The count probabilities for generalized Weibull duration are computed when a = 1, α = 1 and λ = -2, t = 0.25 and t = 1. For the Erlangian count distribution, we compute the probabilities when α = 2, β = 0.8, t = 0.25 and t = 1. A column to indicate the accuracy of our proposed method is included in Table 6.2(b) and Table 6.2(c) since we are comparing against the actual probabilities of the Erlangian count distribution. In all cases, we find that our approximation is accurate up to at least seven decimal places. To illustrate the issue of overflowing which might occur, we present the count probabilities for generalized Weibull duration when a = 2, α = 1 and λ = -2 and t = 1 in Table 6.3. It is clear that in this case, there is a numerical error in the computation of the probabilities when n = 1, 2.

Table 6.2: **Count probabilities for (a) generalized Weibull count distribution, (b) Erlangian count distribution (*t* = 0.25), and (c) Erlangian count distribution (*t* = 1); computation using (i) our proposed method, (ii) formula of the count distribution**

(a) Generalized Weibull count distribution

n	Pr{N(*t*) = *n*} *t* = 0.25		Pr{N(*t*) = *n*} *t* = 1	
	(i)	(ii)	(i)	(ii)
0	0.790123462190233	0.790123456790123	0.444444446077630	0.444444444444444
1	0.185268558281666	0.185268554955749	0.341447772405153	0.341447770099717
2	0.022624019619715	0.022624018469588	0.152421254574663	0.152421252253988
3	0.001862447034136	0.001862446759278	0.047632000079489	0.047631998279757
4	0.000115528824677	0.000115528774610	0.011418307350013	0.011418306220399
5	0.000005746921940	0.000005746914580	0.002217009636005	0.002217009042290
6	0.000000238568216	0.000000238567310	0.000361439244000	0.000361438976100
7	0.000000008496400	0.000000008496304	0.000050759289875	0.000050759184107

(b) Erlangian count distribution (*t* = 0.25)

n	Pr{N(*t*) = *n*}		Accuracy (%)
	(i)	(ii)	
0	0.982476912658251	0.982476903693578	99.99
1	0.017466257275868	0.017466256065664	99.99
2	0.000056765366099	0.000056765332213	99.99
3	0.000000074855777	0.000000074855383	99.99
4	0.000000000053140	0.000000000053138	99.99
5	0.000000000000024	0.000000000000024	100.00
6	0.000000000000000	0.000000000000000	100.00
7	0.000000000000000	0.000000000000000	100.00

(c) Erlangian count distribution (*t* = 1)

n	Pr{N(*t*) = *n*}		Accuracy (%)
	(i)	(ii)	
0	0.808792138560495	0.808792135410999	100.00
1	0.182128011589934	0.182128006788847	99.99
2	0.008895517173780	0.008895515278950	99.99
3	0.000182292662905	0.000182292332810	99.99
4	0.000002035889418	0.000002035857392	99.99
5	0.000000014264304	0.000000014262333	99.99
6	0.000000000068513	0.000000000068429	99.88
7	0.000000000000241	0.000000000000239	99.17

Table 6.2, continued

Table 6.3: **Count probabilities for generalized Weibull count distribution when $a = 2$, $\alpha = 1$ and $\lambda = -2$ and $t = 1$**

n	Pr{N(*t*) = *n*}	
	Formula (6.3.1.1)	Proposed inverse Laplace transform method
0	0.2500	0.2500
1	63.5982	0.2971
2	2.3327	0.2305
3	0.1839	0.1317
4	0.0604	0.0593
5	0.0220	0.0220
6	0.0069	0.0069
7	0.0019	0.0019

The pmf of the inverse Gaussian count distribution has been given in Chapter 5. Chaudhry et al. (2013) computed the probabilities for the inverse Gaussian count distribution but the pmf was not derived. In Table 6.4, we compare the probability function of gamma, inverse Gaussian and Weibull count distributions with those obtained by Chaudhry et al. (2013). The difference in the probabilities is at most two decimal places. In the case of Weibull count distribution, we include only the results when $t = 0.25$ because the algorithm could not converge for $t = 0.60$ and $t = 1$ when $\lambda = 3$, which are the other two values included in Chaudhry et al. (2013). Convergence issues with the Weibull renewal function have been discussed by Constantine and Robinson (1997) whereby they developed a convergent damped exponential series by residue calculations of the Laplace transform of the renewal integral equation for the Weibull renewal function when $\lambda > 1$.

Table 6.4: Probability functions for (a) gamma count distribution, (b) inverse Gaussian count distribution, and (c) Weibull count distribution for selected values of *t*; computation using (i) our proposed method, (ii) method of Chaudhry et al. (2013).

(a) Gamma count distribution

t	Pr{$N(t)$ = 0}		Pr{$N(t)$ = 1}		Pr{$N(t)$ = 2}		Pr{$N(t)$ = 3}		Pr{$N(t)$ = 4}	
	(i)	(ii)	(i)	(ii)	(i)	(ii)	(i)	(ii)	(i)	(ii)
0.1	0.6938	0.6871	0.2341	0.2385	0.0579	0.0602	0.0117	0.0119	0.0021	0.0019
0.4	0.4061	0.4071	0.3092	0.3088	0.1683	0.1677	0.0744	0.0743	0.0283	0.0284
1.25	0.1291	0.1291	0.1952	0.1951	0.2050	0.2050	0.1730	0.1730	0.1249	0.1249

(b) Inverse Gaussian count distribution

t	Pr{$N(t)$ = 0}		Pr{$N(t)$ = 1}		Pr{$N(t)$ = 2}		Pr{$N(t)$ = 3}		Pr{$N(t)$ = 4}	
	(i)	(ii)	(i)	(ii)	(i)	(ii)	(i)	(ii)	(i)	(ii)
0.25	0.7394	0.7445	0.2497	0.2442	0.0108	0.0112	0.0001	0.0001	0.0000	0.0000
0.7	0.3377	0.3390	0.4070	0.4042	0.2044	0.2062	0.0460	0.0457	0.0047	0.0046
1.0	0.1623	0.1623	0.2865	0.2869	0.2871	0.2867	0.1763	0.1762	0.0681	0.0683

(c) Weibull count distribution

t	Pr{$N(t)$ = 0}		Pr{$N(t)$ = 1}		Pr{$N(t)$ = 2}		Pr{$N(t)$ = 3}		Pr{$N(t)$ = 4}	
	(i)	(ii)	(i)	(ii)	(i)	(ii)	(i)	(ii)	(i)	(ii)
0.25	0.9845	0.9841	0.0155	0.0159	0.0000	0.0000	0.0000	-	0.0000	-

6.4 Renewal function and variance

Using the probability of the counts, we also computed the renewal function and variance function for comparison with those obtained by Baxter et al. (1981) and Chaudhry et al. (2013). The details are presented in Table 6.5. In most cases, the values computed using our proposed method is closer to that of Baxter et al. (1981). Baxter et al. (1981) has verified the accuracy of their extended cubic splining algorithm through comparisons with previous tabulations for the Weibull count distribution in the literature and a direct evaluation of the incomplete gamma integral for the gamma count distribution.

Table 6.5: Renewal and variance function for (a) gamma count distribution, (b) inverse Gaussian count distribution, and (c) Weibull count distribution for selected values of *t*; computation using (i) our proposed method, (ii) method of Baxter et al. (1981), and (iii) method of Chaudhry et al. (2013).

(a) Gamma count distribution

t	Renewal Function			Variance Function		
	(i)	(ii)	(iii)	(i)	(ii)	(iii)
0.1	0.3953	0.3933	0.4040	0.4580	0.4485	0.4623
0.4	1.0560	1.0550	1.0545	1.3954	1.3901	1.3970
1.25	2.6662	2.6653	2.6663	4.0491	4.0441	4.0487

(b) Inverse Gaussian count distribution

t	Renewal Function			Variance Function		
	(i)	(ii)	(iii)	(i)	(ii)	(iii)
0.25	0.2716	0.2715	0.2669	0.2198	0.2200	0.2188
0.7	0.9739	0.9739	0.9736	0.7717	0.7718	0.7732
1.0	1.7636	1.7638	1.7635	1.5290	1.5293	1.5294

(c) Weibull count distribution

t	Renewal Function			Variance Function		
	(i)	(ii)	(iii)	(i)	(ii)	(iii)
0.25	0.0155	0.0156	0.0159	0.0153	0.0154	0.0156

6.5 Applications of the Count Distributions

In this section, the pmf of the count distributions are evaluated using the numerical inverse Laplace transform method discussed in the preceding sections. Maximum likelihood (ML) estimation of the parameters is performed via numerical global optimization using the simulated annealing (Goffe, Ferrier & Rogers, 1994) algorithm. Although we have derived the count distribution for a time period of length *t*, we consider a unit time interval $t = 1$ without loss of any generality. In this section we follow Cochran's (1954) recommendation to group cells with expected frequencies of less than 1.0.

6.5.1 Over Dispersed Data

We fit count distributions that are able to model over dispersion to two over dispersed data sets. Consequently, the inverse Gaussian count distribution and the hypoexponential count distribution are not discussed in this section, since they are able to model only under dispersed data sets.

Table 6.6 gives the observed and expected frequency table on the number of doctor consultations in a two-week period from the 1977-78 Australian Health Surveys (Cameron & Trivedi, 1986). Such data is useful in health economics research which studies the link between health care utilization and economic variables. The data set is over dispersed with a mean of 0.3017 and variance 0.6370. Based on the AIC value and chi-square goodness-of-fit test statistic, the Weibull count distribution gives the best fit for this data set.

Table 6.6: Number of consultations with specialists or doctors in a two-week period (Cameron & Trivedi, 1986)

Count	Obs	Exp	G	Hyper Exp	Conv G	W	GW
0	4141	3838.18	3985.02	3986.99	3986.08	3986.81	3986.61
1	782	1158.11	926.48	924.16	925.02	924.21	924.23
2	174	174.72	214.40	214.21	214.37	214.28	214.37
3	30	17.57	49.40	49.65	49.61	46.69	49.74
4	24	1.33	11.34	11.51	11.47	11.52	11.55
5	9	0.08	3.36	3.47	3.44	3.48	3.49
6	12						
7	12						
8	5						
9	1						
Total		5190.00	5190.00	5190.00	5190.00	5190.00	5190.00
Number of classes		6	6	6	6	6	6
d.f.		4	3	2	1	3	2
χ^2		18540.47	436.57	420.16	424.69	419.14	417.46
p-value		< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Log-likelihood function value		-3983.19	-3660.56	-3657.93	-3658.75	-3657.83	-3657.83
AIC		7968.39	7325.11	7321.87	7325.50	7319.66	7321.27

Table 6.7 presents the count frequency distribution for the labour mobility data (Winkelmann & Zimmermann, 1995), which is a topic of interest in labour economics. The data has a sample mean of 0.7906 and variance 1.6050. For this data set, the generalized Weibull count distribution gives the best fit.

For the over dispersed data sets, the count model with hyperexponential duration provides a closer fit than the exponential and gamma duration models in terms of AIC value. The Weibull count distribution and the generalized Weibull count distribution give the best fit to the data in Table 6.6 and Table 6.7, respectively. Since the Weibull and generalized Weibull distribution does not have a closed form Laplace transform, the model fitting takes up a significantly longer time. In the case of distributions with closed Laplace transform, the hyperexponential count distribution gives the best fit for both over dispersed data sets presented in this section. As such, when computational time is a concern, the hyperexponential count distribution may serve as a feasible model to fit these data sets.

Table 6.7: Labour Mobility (Winkelmann and Zimmermann, 1995)

Count	Obs	Exp	G	Hyper Exp	Conv G	W	GW
0	465	366.04	437.51	450.68	450.40	451.05	457.51
1	183	289.38	207.75	198.99	199.19	198.66	197.80
2	89	114.39	93.49	87.86	88.00	87.65	84.08
3	39	30.15	40.33	38.79	38.83	38.74	35.70
4	17	5.96	16.80	17.13	17.12	17.16	15.66
5	5	1.08	6.80	7.56	7.54	7.61	7.39
6	1		2.68	3.34	3.32	3.39	3.83
7	6		1.03	1.47	1.46	1.51	2.15
8	0		0.62	1.16	1.14	1.22	1.25
9	1						1.62
10	1						
Total	807						
Number of classes		6	9	9	9	9	10
d.f.		4	6	5	4	6	6
χ^2		248.84	33.47	18.75	19.06	18.14	13.05
p-value		< 0.001	< 0.001	0.0021	0.0008	0.0059	0.0423
Log-likelihood function value		-1083.09	-995.35	-991.69	-991.62	-991.55	-990.07
AIC		2168.18	1992.71	1985.38	1991.24	1987.10	1986.15

We also verify that the convolution of two exponentials count distribution gives exactly the same fit as the simple Poisson distribution, implying that this distribution is not suitable for over dispersed count data. The inverse Gaussian count distribution also gives a poor fit to both of these data sets. This coincides with the characteristic of inter-arrival time distributions which has an increasing hazard function.

6.5.2 Under Dispersed Data

Table 6.8 gives the distribution on completed fertility in a group of Swedish women (Melkersson & Rooth, 2000). The data set is slightly under dispersed with mean 2.1641 and variance 1.7814. From the chi-square goodness-of-fit value and the AIC value, the convolution of two gamma count distribution is the best fitting model for this data set.

Table 6.8: Completed fertility in a group of Swedish women (Melkersson & Rooth, 2000)

Count	Obs	Exp	G	Conv Exp	Conv G	IG	W	GW
0	114	134.38	96.43	100.90	104.88	114.80	94.47	134.03
1	205	290.81	292.65	285.19	273.29	266.41	298.08	290.94
2	466	314.67	354.96	352.34	357.09	342.10	354.52	314.80
3	242	226.99	249.85	254.80	260.57	266.72	246.32	227.04
4	85	122.81	119.26	122.08	121.96	130.34	118.09	122.83
5	35	53.15	42.08	41.63	40.12	40.47	42.65	53.16
6	16	19.17	11.60	10.60	9.87	8.04	12.26	19.18
7	4	5.93	3.17	2.46	2.23	1.11	3.61	5.93
8 - 12	3	2.09						2.09
Total		1170.00	1170.00	1170.00	1170.00	1170.00	1170.0	1170.0
Number of		10	9	9	9	9	9	10
d.f.		8	6	6	4	4	6	6
χ^2		121.57	81.78	84.99	78.26	116.84	83.20	121.40
p-value		< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Log-likelihood function value		-1956.66	-1942.17	-1944.51	-1940.25	-1958.77	-1942.58	-1956.59
AIC		3915.31	3888.33	3893.03	3888.50	3921.54	3889.16	3919.18

Skellam's (1948) data on secondary association of chromosomes in Brassika is given in Table 6.9. The data is under dispersed with a mean of 1.7418 and variance 0.8562. Based on the chi-square values, the convolution of two gamma count distribution gives the best fit to this data set. However, it is to be noted that the small number of classes affects the distribution of the chi-square statistic since its degree of freedom depends on the number of parameters (i.e. four in the convolution of two gamma count distribution). In this event, the two-parameter inverse Gaussian count distribution gives the best fit with a chi-square value of 21.03.

Table 6.9: Secondary association of chromosomes in Brassika (Skellam, 1948)

Count	Observed	Exp	G	Conv Exp	Conv G	IG	W	GW
0	32	114.96	23.77	31.35	36.74	26.57	24.55	56.15
1	103	123.64	118.70	116.02	90.21	109.58	123.14	104.99
2	122	66.49	128.86	118.25	133.85	134.90	122.16	90.14
3	80	23.83	53.35	54.54	75.79	56.96	52.11	52.13
> 4	0	8.07	12.31	16.84	0.41	8.98	15.04	33.60
Total	337	337.00	337.00	337.00	337.00	337.00	337.00	337.00
Number of classes		5	5	5	5	5	5	5
d.f.		3	2	2	0	2	2	1
χ^2		250.11	30.91	30.32	3.66	21.03	35.53	70.18
p-value		< 0.001	< 0.001	< 0.001	-	< 0.001	< 0.001	< 0.001
Log-likelihood function value		-489.88	-457.06	-459.29	-438.68	-450.91	-460.66	-487.66
AIC		981.76	918.13	922.58	885.37	905.81	925.32	981.32

The hyperexponential count distribution gives a very poor fit to both data sets. This observation coincides with the characteristic of the hyperexponential inter-arrival times' distribution which has a decreasing hazard function thus it is suitable for modelling over dispersed data sets.

The ML estimates are given in Table 6.10. The standard errors of the ML estimates for the distribution which gives the best fit to the respective data sets are computed through bootstrapping and given in the same table. For numerical stability, we transform the parameters for the generalized Weibull count distributions to its corresponding reciprocals prior to performing ML estimation. For the data on monthly doctor consultations and labour mobility, it is noted that both the ML estimates of β for the gamma duration model and λ_1 for the hyperexponential duration model approaches the boundary of 0.

Table 6.10: ML estimates of the fitted distributions

Data set	**ML estimates of parameters (standard error)**
Legend	Exp = Exponential; G = Gamma; HyperExp = Hyperexponential; ConvExp = Convolution of two exponentials; ConvG = Convolution of two gamma; IG = Inverse Gaussian; W = Weibull; GW = Generalized Weibull
Table 6.6 (Number of consultations with doctors or specialists)	$\hat{\lambda}$ = 0.3017 [(Exp] $\hat{\alpha}$ = 0.0545, $\hat{\beta}$ = 0.0000 [G] $\hat{\lambda}_1$ = 0.0000, $\hat{\lambda}_2$ = 43.8787, $\hat{p}$ = 0.7682 [HyperExp] $\hat{\alpha}_1$ = 0.0293, $\hat{\beta}_1$ = 0.0000, $\hat{\alpha}_2$ = 0.0000, $\hat{\beta}_2$ = 7.7068 [ConvG] $\hat{\alpha}$ = 16.6245 (1.8436), $\hat{\lambda}$ = 7.4567 (0.2119) [W] $\hat{a}$ = 3090.40, $\hat{\alpha}$ = 0.3673, $\hat{\lambda}$ = -0.2514 [GW]
Table 6.7 (Labour mobility data)	$\hat{\lambda}$ = 0.7905 [Exp] $\hat{\alpha}$ = 0.01, $\hat{\beta}$ = 0.1870 [G] $\hat{\lambda}_1$ = 0.0000, $\hat{\lambda}_2$ = 19.4552, $\hat{p}$ = 0.5585 [HyperExp] $\hat{\alpha}_1$ = 0.0282, $\hat{\beta}_1$ = 0.0000, $\hat{\alpha}_2$ = 0.0000, $\hat{\beta}_2$ = 67.0468 [ConvG] $\hat{\alpha}$ = 0.0551, $\hat{\lambda}$ = 3.7724 [W] $\hat{a}$ = 128.6581 (74.4390), $\hat{\alpha}$ = 1.2437 (0.1623), $\hat{\lambda}$ = -0.1208 (0.0281) [GW]
Table 6.8 (Completed fertility in Swedish women)	$\hat{\lambda}$ = 2.1641 [Exp] $\hat{\alpha}$ = 1.3532, $\hat{\beta}$ = 3.1046 [G] β_1 = 2.6091, β_2 = 17.8015 [ConvExp] $\hat{\alpha}_1$ = 1.6063 (0.8888), $\hat{\beta}_1$ = 4.0972 (1.6213), $\hat{\alpha}_2$ = 0.0008 (0.2161), $\hat{\beta}_2$ = 0.0000 (2.096) [ConvG] $\hat{\lambda}$ =0.4487, $\hat{\mu}$ = 0.4602 [IG] $\hat{\alpha}$ = 2.1692, $\hat{\lambda}$ = 0.8462 [W] $\hat{a}$ = 2.1642, $\hat{\alpha}$ = 1.0000, $\hat{\lambda} \to \infty$ [GW]
Table 6.9 (Secondary association in chromosomes)	$\hat{\lambda}$ = 1.0755 [Exp] $\hat{\alpha}$ = 2.4717, $\hat{\beta}$ = 5.0429 [G] β_1 = 3.9803, β_2 = 3.9804 [ConvExp] $\hat{\alpha}_1$ = 89.9209, $\hat{\beta}_1$ = 336.8173, $\hat{\alpha}_2$ = 0.3218, $\hat{\beta}_2$ = 1.2048 [ConvG] $\hat{\lambda}$ = 1.0828 (0.0932), $\hat{\mu}$ = 0.4970 (0.0158) [IG] $\hat{\alpha}$ = 1.8249, $\hat{\lambda}$ = 0.6258 [W] $\hat{a}$ = 1.7424, $\hat{\alpha}$ = 1.0000, $\hat{\lambda} \to \infty$ [GW]

6.6 Hypothesis Testing for the Hyperexponential Count Distribution

From the results in the previous section, the hyperexponential count distribution can serve as an alternative to the simple Poisson distribution or gamma distribution to model over dispersed data. To determine the significance of the phase parameter p in the hyperexponential count distribution, we perform the likelihood ratio test to test the hypotheses H_0: $p = 0$ (exponential duration) versus H_1: $p \neq 0$ (hyperexponential duration).

For all the over dispersed data sets, the likelihood ratio test statistic has a value of greater than 70.0 leading to the rejection of H_0 at a significance level of 0.05.

6.7 Conclusion

In this chapter, an efficient and accurate method to compute the probabilities of a count distribution arising from non-exponential inter-arrival times between event counts is proposed and discussed. The implementation of this numerical inverse Laplace transform technique is straightforward when the Laplace transform of the inter-arrival times' distribution is available in closed form. This method removes the computational hurdle in applying a wide range of lifetime distributions for modelling the inter-arrival times between events. When the Laplace transform of the inter-arrival time distribution is not available in closed form, other methods to approximate the Laplace transform for numerical inversion can be explored, such as the infinite series, Gaussian quadrature, Laguerre method and the continued fractions technique.

CHAPTER 7: CONCLUSION AND FURTHER WORK

In this thesis, some families of count distributions for modelling over dispersion, under dispersion and zero-inflation have been proposed and studied. In a mixed Poisson distribution, the choice of the mixing distribution is crucial especially in determining the amount of zero-inflation and tail length of the resulting Poisson mixture. The extended generalized inverse Gaussian (EGIG) and exponentiated Weibull distributions have been earmarked as mixing distributions in part of the work in this thesis due to their shape flexibility and desirable statistical properties such as regularity. Two new mixed Poisson distributions, namely the generalized Sichel distribution and Poisson-exponentiated Weibull distribution have been shown to model over dispersion, zero-inflation and long-tailed data very well. Issues related to the computational hurdle posed by an intractable probability mass function have been addressed through a Monte Carlo simulation technique and an EM-type algorithm for parameter estimation.

We have also studied a family of count distributions arising from non-exponential duration in a renewal process. The duration models considered are the inverse Gaussian distribution, convolution of two gamma distributions and a two-component exponential mixture distribution. This family of count distributions are able to model over dispersion and under dispersion and serves as a satisfactory, if not better alternative to existing models in the same context. Given the distribution of the counts, this approach provides an insight to the waiting times or inter-arrival times in a stochastic process, and vice versa.

A numerical inverse Laplace transform technique is proposed to facilitate computation of count probabilities arising from non-exponential duration models in a stochastic process. These count probabilities often involve special mathematical functions and infinite series. In some cases, the probability mass function is expressed as multiple integrals. In the past, such computation requires extensive numerical methods and may result in an inevitable numerical overflow. The numerical inverse Laplace transform technique is able to avoid these issues and provide sufficiently accurate results.

Future work from the findings of this thesis will be to extend the distributions in a regression modelling context. When such information is available, inclusion of covariates in modelling the count data will give further insight to the analyst to answer the research question of interest. In particular, Rigby, Stasinopoulos and Akantziliotou (2008) have developed a generalized additive model for location, shape and scale and this approach could be further explored with the generalized Sichel distribution.

A relatively simple but flexible distribution known as the exponentiated Nadarajah Haghighi distribution (Lemonte, 2013) can be considered as the underlying duration model for a new count distribution in a stochastic process. The numerical inverse Laplace transform technique for computing probabilities is straightforward when the Laplace transform of the duration distribution is in closed form. Further work can be done to examine approximations to the Laplace transform when it is not in closed form and its implications on the numerical inverse Laplace technique for computing probabilities, in terms of accuracy and efficiency.

REFERENCES

Abate, J., & Valkó, P. P. (2004). Multi-precision Laplace transform inversion. *International Journal for Numerical Methods in Engineering*, *60*(5), 979–993.

Abate, J., & Whitt, W. (1992). The Fourier-series method for inverting transforms of probability distributions. *Queueing Systems*, *10*(1-2), 5–87.

Abate, J., & Whitt, W. (1995). Numerical inversion of Laplace transforms of probability distributions. *ORSA Journal on Computing*, *7*(1), 36–43.

Aitkin, M. (1996). A general maximum likelihood analysis of overdispersion in generalised linear models. *Statistics and Computing*, *6*, 251–262.

Aitkin, M. (1999). A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics*, *55*(1), 117–128.

Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, *19*(6), 716–723.

Arbous, A. G., & Sichel, H. S. (1954). New techniques for the analysis of absenteeism data. *Biometrika*, *41*, 77–90.

Atkinson, A. C., & Yeh, L. (1982). Inference for Sichel's compound Poisson distribution. *Journal of the American Statistical Association*, *77*(377), 153–158.

Balakrishnan, N., & Pal, S. (2012). EM algorithm-based likelihood estimation for some cure rate models. *Journal of Statistical Theory and Practice*, *6*(4), 698–724.

Banerjee, A. K., & Bhattacharyya, G. K. (1976). A purchase incidence model with Inverse Gaussian interpurchase time. *Journal of the American Statistical Association*, *71*(356), 823–829.

Bardwell, G. E., & Crow, E. L. (1964). A two-parameter family of hyper-Poisson distributions. *Journal of the American Statistical Association.*, *59*, 133–141.

Baxter, L. A., Scheuer, E. M., Blischke, W. R., & McConalogue, D. J. (1981). *Renewal Tables: Tables of Functions Arising in Renewal Theory*.

Bohning, D., Dietz, E., Schlattmann, P., Mendonca, L., & Kirchner, U. (1999). The zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, *162*(2), 195–209.

Boucher, J.-P., & Guillén, M. (2009). A survey on models for panel count data with applications to insurance. *Revista de La Real Academia de Ciencias Exactas, Fisicas Y Naturales. Serie A. Matematicas*, *103*(2), 277–294.

Broek, J. van den. (1995). A score test for zero inflation in a Poisson distribution. *Biometrics*, *51*(2), 738–743.

Brooks, R.J. (1984). Approximate likelihood ratio tests in the analysis of beta-binomial data. *Applied Statistics*, *33*, 285-289.

Bulmer, M. G. (1974). On fitting the Poisson lognormal distribution to species-abundance data. *Biometrics*, *30*(1), 101–110.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*, *33*(2), 261–304.

Burnham, K. P., & Anderson, D. R. (2002). *Model Selection and Multimodel Inference* (2nd ed.). New York: Springer-Verlag.

Buse, A. (1982). The likelihood ratio, Wald, and Lagrange multiplier tests: An expository note. *The American Statistician*, *36*(3), 153–157.

Cameron, A. C., & Johansson, P. (1997). Count data regression using series expansion: With applications. *Journal of Applied Econometrics*, *12*(3), 203–223.

Cameron, A. C., & Trivedi, P. K. (1986). Econometric models based on count data: Comparisons and applications of some estimators and tests. *Journal of Applied Econometrics*, *1*(1), 29–54.

Carroll, R. J., & Ruppert, D. (1988). *Transformation and Weighting in Regression*. London: Chapman & Hall.

Casella, G., & Berger, R. L. (2002). *Statistical Inference*. Duxbury Press.

Cassie, R. M. (1962). Frequency distribution models in the ecology of plankton and other organisms. *Journal of Animal Ecology*, *31*(1), 65–92.

Castillo, J., & Pérez-Casany, M. (2005). Overdispersed and underdispersed Poisson generalizations. *Journal of Statistical Planning and Inference*, *134*, 486–500.

Celeux, G., & Diebolt, J. (1985). The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Computational Statistics Quarterly*, *2*, 73-82.

Celeux, G., Chauveau, D., & Diebolt, J. (1995). *On Stochastic Versions of the EM Algorithm*. Retrieved from https://hal.inria.fr/inria-00074164.

Chan, K.S., & Ledolter, J. (1995). Monte Carlo EM estimation for time series models involving counts. *Journal of the American Statistical Association*, *90*(429), 242-252.

Chaudhry, M. L., Yang, X., & Ong, B. (2013). Computing the distribution function of the number of renewals. *American Journal of Operations Research*, *3*, 380–386.

Cheng, L., Geedipally, S. R., & Lord, D. (2013). The Poisson-Weibull generalized linear model for analyzing motor vehicle crash data. *Safety Science*, *54*, 38–42.

Chhikara, R. S., & Folks, J. L. (1977). The inverse Gaussian distribution as a lifetime model. *Technometrics*, *19*(4), 461–468.

Choudhury, A. (2005). A simple derivation of moments of the exponentiated Weibull distribution. *Metrika*, *62*, 17–22.

Chung, Y., & Lindsay, B.G. (2015). Convergence of the EM algorithm for continuous mixing distributions. *Statistics and Probability Letters*, *96*, 190–195.

Cochran, W. G. (1954). Some methods for strengthening the common chi-square tests. *Biometrics*, *10*(4), 417–451.

Constantine, A. G., & Robinson, N. I. (1997). The Weibull renewal function for moderate to large arguments. *Computational Statistics & Data Analysis*, *24*(1), 9–27.

Consul, P. C. (1989). *Generalized Poisson Distributions*. New York: Dekker.

Consul, P. C., & Jain, G. C. (1973). A generalization of the Poisson distribution. *Technometrics*, *15*(4), 791–799.

Conway, R. W., & Maxwell, W. L. (1962). A queuing model with state dependent service rates. *Journal of Industrial Engineering*, *12*, 132–136.

Corana, A., Marchesi M., Martini, C., & Ridella, S. (1987). Minimizing multimodal functions of continuous variables with the “simulated annealing” algorithm. *ACM Transactions on Mathematical Software*, *13*(3), 262–280.

Cox, D. R. (1962). *Renewal Theory*. London: Methuen.

Cox, D. R. (1983). Some Remarks on Overdispersion. *Biometrika*, *70*(1), 269–274.

Cox, D. R., & Hinkley, D. V. (1974). *Theoretical Statistics*. London: Chapman & Hall/CRC.

Dalrymple, M. L., Hudson, I. L., & Ford, R. P. K. (2003). Finite mixture, zero-inflated Poisson and hurdle models with application to SIDS. *Computational Statistics & Data Analysis*, *41*(3-4), 491–504.

Dean, C., Lawless, J. F., & Willmot, G. E. (1989). A mixed Poisson-inverse Gaussian regression model. *Canadian Journal of Statistics*, *17*, 171–182.

Dempster, A. P. A., Laird, N. M. N., & Rubin, D. D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B Methodological*, *39*(1), 1–38.

Dubner, H., & Abate, J. (1968). Numerical inversion of Laplace transforms by relating them to the finite Fourier cosine transform. *Journal of the ACM*, *15*(1), 115–123.

Eaton, W. W., & Whitmore, G. A. (1977). Length of stay as a stochastic process: A general approach and application to hospitalization for schizophrenia. *The Journal of Mathematical Sociology*, *5*(2), 273–292.

Efron, B. (1981). Nonparametric standard errors and confidence intervals. *The Canadian Journal of Statistics*, *9*(2), 139–158.

Efron, B. (1986). Exponential families and their use in generalized linear regression. *Journal of the American Statistical Association*, *81*(395), 709–721.

Efron, B., & Hinkley, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: observed versus expected Fisher information. *Biometrika*, *65*(3), 457–482.

Erdelyi, A. M., Oberhettinger, F., & Tricomi, F. G. (1953). *Higher Transcendental Functions*. New York: McGraw-Hill.

Famoye, F., & Singh, K. P. (2006). Zero-inflated generalized Poisson regression model with an application to domestic violence data. *Journal of Data Science*, *4*, 117–130.

Fang, K.-T., & Wang, Y. (1994). *Number-theoretic methods in statistics*. London: Chapman & Hall/CRC.

Feldmann, A., & Whitt, W. (1998). Fitting mixtures of exponentials to long-tail distributions to analyze network performance models. *Performance Evaluation*, *31*, 245–279.

Feller, W. (1943). On a general class of “contagious” distributions. *Annals of Mathematical Statistics*, *14*, 389–400.

Fletcher, R. (1987). *Practical methods of optimization*. Chichester: John Wiley & Sons.

Freedman, D. A. (2007). How can the score test be inconsistent? *The American Statistician*, *61*(4), 291–295.

From, S. G. (2004). Approximating the distribution of a renewal process using generalized Poisson distributions. *Journal of Statistical Computation and Simulation*, *74*(9), 667–681.

Gathy, M., & Lefèvre, C. (2010). On the Lagrangian Katz family of distributions as a claim frequency model. *Insurance: Mathematics and Economics*, *47*, 76–83.

Ghitany, M. E., & Al-Mutairi, D. K. (2009). Estimation methods for the discrete Poisson-Lindley distribution. *Journal of Statistical Computation and Simulation*, *79*(1), 1–9.

Goffe, W. L., Ferrier, G. D., & Rogers, J. (1994). Global optimization of statistical functions with simulated annealing. *Journal of Econometrics*, *60*, 65–99.

Gómez-Déniz, E., Sarabia, J. M., & Calderín-Ojeda. (2008). Univariate and multivariate versions of the negative binomial-inverse Gaussian distributions with applications. *Insurance: Mathematics and Economics*, *42*, 39–49.

Gómez-Déniz, E., Sarabia, J. M., & Calderín-Ojeda, E. (2011). A new discrete distribution with actuarial applications. *Insurance: Mathematics and Economics*, *48*(3), 406–412.

Grandell, J. (1997). *Mixed Poisson Processes*. New York: Chapman & Hall/CRC.

Greenwood, M., & Yule, G. K. (1920). An enquiry into the nature of the frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or repeated accidents. *Journal of Royal Statistical Society. Series A*, *83*, 255–279.

Grundy, P. M. (1951). The expected frequencies in a sample of an animal population in which the abundances of species are log-normally distributed. Part I. *Biometrika*, *38*(3), 427–434.

Gupta, P. L., Gupta, R. C., & Ong, S. H. (2004). Modelling count data by random effect Poisson model. *Sankhyā: The Indian Journal of Statistics*, *66*(3), 548–565.

Gupta, P. L., Gupta, R. C., & Tripathi, R. C. (1996). Analysis of zero-adjusted count data. *Computational Statistics and Data Analysis*, *23*(2), 207–218.

Gupta, P. L., Gupta, R. C., & Tripathi, R. C. (2005). Score test for zero inflated generalized Poisson regression model. *Communications in Statistics - Theory and Methods*, *33*(1), 47–64.

Gupta, R. C., & Ong, S. H. (2004). A new generalization of the negative binomial distribution. *Computational Statistics and Data Analysis*, *45*(2), 287–300.

Gupta, R. C., & Ong, S. H. (2005). Analysis of long-tailed count data by Poisson mixtures. *Communications in Statistics - Theory and Methods*, *34*(3), 557–573.

Gupta, R. C., & Viles, W. (2011). Statistical inference for the extended generalized inverse Gaussian model. *Journal of Statistical Computation and Simulation*, 1–18.

Gurmu, S., & Trivedi, P. K. (1996). Excess zeros in count models for recreational trips. *Journal of Business & Economic Statistics*, *14*(4), 469–477.

Hall, D. B. (2000). Zero-inflated poisson and binomial regression with random effects: A Case Study. *Biometrics*, *56*, 1030–1039.

He, H., Tang, W., Wang, W., & Crits-Christoph, P. (2014). Structural zeroes and zero-inflated models. *Shanghai Archives of Psychiatry*, *26*(4), 236–242.

Heilbron, D. C. (1994). Zero-altered and other regression models for count data with added zeros. *Biometrical Journal*, *36*, 531–547.

Holgate, P. (1970). The modality of some compound Poisson distributions. *Biometrika*, *57*, 666–667.

Holla, M. S. (1967). On a Poisson-inverse Gaussian distribution. *Metrika*, *11*(1), 115–121.

Holla, M. S., & Bhattacharya, S. K. (1965). On a discrete compound distribution. *Annals of the Institute of Statistical Mathematics*, *15*, 377–384.

Hougaard, P., Lee, M. L., & Whitmore, G. A. (1997). Analysis of overdispersed count data by mixtures of Poisson variables and Poisson processes. *Biometrics*, *53*(4), 1225–1238.

Hurvich, C. M., & Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, *76*(2), 297–307.

Ickowicz, A., & Sparks, R. (2015). Modelling hospital length of stay using convolutive mixtures distributions. *arXiv preprint arXiv:1512.02688*.

Irwin, J. O. (1968). The generalized Waring distribution applied to accident theory. *Journal of the Royal Statistical Society. Series A (General)*, *131*(2), 205–225.

Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., & Mann, M. (2005). Exponentially Modified Protein Abundance Index (emPAI) for Estimation of Absolute Protein Amount in Proteomics by the Number of Sequenced Peptides per Protein. *Molecular & Cellular Proteomics*, *4*(9), 1265–1272.

Izsák, R. (2008). Maximum likelihood fitting of the Poisson lognormal distribution. *Environmental and Ecological Statistics*, *15*(2), 143–156.

Jain, G. C., & Consul, P. C. (1971). A generalized negative binomial distribution. *SIAM Journal on Applied Mathematics*, *21*(4), 501–513.

Jank, W. (2005). Quasi-Monte Carlo sampling to improve the efficiency of Monte Carlo EM. *Computational Statistics and Data Analysis*, *48*(4), 685-701.

Jansakul, N., & Hinde, J. (2009). Score tests for extra-zero models in zero-inflated negative binomial models. *Communications in Statistics - Simulation and Computation*, *38*(1), 92–108.

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1994). *Continuous Univariate Distributions, Vol 1* (2nd ed.). New York: John Wiley & Sons.

Johnson, N. L., Kemp, A. W., & Kotz, S. (2005). *Univariate Discrete Distributions* (3rd ed.). New York: John Wiley and Sons.

Jørgensen, B. (1982). *Statistical properties of the generalized inverse Gaussian distribution. Lecture Notes in Statistics (Vol. 9)*. New York: Springer Verlag.

Jørgensen, B. (1987). Exponential dispersion models. *Journal of the Royal Statistical Society. Series B (Methodological)1*, *49*(2), 127–162.

Jørgensen, B., & Kokonendji, C. C. (2016). Discrete dispersion models and their Tweedie asymptotics. *AStA Advances in Statistical Analysis*, *100*(1), 43–78.

Jose, K. K., & Abraham, B. (2011). A count model based on Mittag-Leffler interarrival times. *STATISTICA, LXXI*(4), 501–514.

Jose, K. K., & Abraham, B. (2013). A Counting process with Gumbel inter-arrival times for modeling climate data. *Journal of Environmental Statistics, 4*(5), 13.

Kadane, J. B., Krishnan, R., & Shmueli, G. (2006). A data disclosure policy for count data based on the COM-Poisson distribution. *Management Science*, *52*(10), 1610–1617.

Karlis, D. (2001). A general EM approach for maximum likelihood estimation in mixed Poisson regression models. *Statistical Modelling*, *1*, 305–318.

Karlis, D. (2005). EM algorithm for mixed Poisson and other discrete distributions. *ASTIN Bulletin*, *35*(1), 3–24.

Karlis, D., & Xekalaki, E. (2003). Choosing initial values for the EM algorithm for finite mixtures. *Computational Statistics and Data Analysis*, *41*(3-4), 577–590.

Karlis, D., & Xekalaki, E. (2005). Mixed Poisson distributions. *International Statistical Review*, *73*(1), 35–58.

Kemp, A. W., & Kemp, C. D. (1966). An alternative derivation of the Hermite distribution. *Biometrika*, *53*, 627–628.

Kirkpatrick, S., Gelatt, C.D., J., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, *220*(4598), 671–680.

Kokonendji, C. C. (2014). Over- and Underdispersion Models. In N. Balakrishnan (Ed.), *Methods and Applications of Statistics in Clinical Trials: Planning, Analysis, and Inferential Methods* (pp. 506–526). New Jersey: John Wiley & Sons.

Kokonendji, C. C., & Khoudar, M. (2004). On strict arcsine distribution. *Communications in Statistics - Theory and Methods*, *33*, 993–1006.

Koziol, J., Griffin, N., Long, F., Li, Y., Latterich, M., & Schnitzer, J. (2013). On protein abundance distributions in complex mixtures. *Proteome Science*, *11*(1).

Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, *34*(1), 1–14.

Lancaster, T. (1972). A stochastic model for the duration of a strike. *Journal of the Royal Statistical Society. Series A (General)*, *135*(2), 257–271.

Lange, K. (1995). A quasi-Newton acceleration of the EM algorithm. *Statistica Sinica*, *5*, 1-18.

Lawless, J. F. (1987). Regression methods for Poisson process data. *Journal of the American Statistical Association*, *82*(399), 808–815.

Lee, L. (1996). Specification and estimation of count data regression and sample selection models - a counting process and waiting time approach. Retrieved from http://repository.ust.hk/ir/Record/1783.1-2562.

Lemonte, A. (2013). A new exponential-type distribution with constant, decreasing, increasing, upside-down bathtub and bathtub-shaped failure rate function. *Computational Statistics and Data Analysis*, *62*, 149-170.

Letac, G., & Mora, M. (1990). Natural real exponential families with cubic variance functions. *The Annals of Statistics*, *18*(1), 1-37.

Li, C.-S., Lu, J.-C., Park, J., Kim, K., Brinkley, P. A., & Peterson, J. P. (1999). Multivariate zero-inflated Poisson models and their applications. *Technometrics*, *41*(1), 29–38.

Lindsey, J. K. (1998). Counts and times to events. *Statistics in Medicine*, *17*(15-16), 1745–51.

Lindsey, J. K. (1999). On the use of corrections for overdispersion. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *48*(4), 553–561.

Louis, T. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B*, *44*, 226-233.

Lord, D., Guikema, S. D., & Geedipally, S. R. (2008). Application of the Conway-Maxwell-Poisson generalized linear model for analyzing motor vehicle crashes. *Accident Analysis and Prevention*, *40*(3), 1123–1134.

Mahmoudi, E., & Zakerzadeh, H. (2010). Generalized Poisson-Lindley distribution. *Communications in Statistics - Theory and Methods*, *39*(10), 1785–1798.

Melkersson, M., & Rooth, D.-O. (2000). Modeling female fertility using inflated count data models. *Journal of Population Economics*, *13*(2), 189–203.

McLachlan, G. J. (1997). On the EM algorithm for overdispersed count data. *Statistical Methods in Medical Research*, *6*(1), 76-98.

McShane, B., Adrian, M., Bradlow, E. T., & Fader, P. S. (2008). Count models based on Weibull interarrival times. *Journal of Business & Economic Statistics*, *26*(3), 369–378.

Meng, X.-L., & Rubin, D.B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika*, *80*(2), 267–278.

Moore, D.F. (1986). Asymptotic properties of moment estimates for overdispersed counts and proportions. *Biometrika*, *73*, 583-588.

Morgan, B. J. T., Palmer, K. J., & Ridout, M. S. (2007). Negative score test statistic. *The American Statistician*, *61*(4), 285–288.

Morris, C. N. (1982). Natural exponential families with quadratic variance functions. *The Annals of Statistics*, *10*(1), 65–80.

Moschopoulos, P. G. (1985). The distribution of the sum of independent gamma random variables. *Annals of the Institute of Statistical Mathematics*, *37*(1), 541–544.

Mudholkar, G. S., Srivastava, D. K., & Freimer, M. (1995). The exponentiated Weibull family: A reanalysis of the bus-motor-failure data. *Technometrics*, *37*(4), 436–445.

Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of Econometrics*, *33*(3), 341–365.

Mullahy, J. (1997). Heterogeneity, excess zeros, and the structure of count data models. *Journal of Applied Econometrics*, *12*(3), 337–350.

Nair, M. T., & Abdul, E. S. (2010). Finite mixture of exponential model and its applications to renewal and reliability theory. *Journal of Statistical Theory and Practice*, *4*(3), 367–373.

Nelder, J. A., & Pregibon, D. (1987). An extended quasi-likelihood function. *Biometrika*, *74*(2), 221–232.

Neyman, J., & Pearson, E. S. (1928). On the use and interpretation of certain test criteria. *Biometrika*, *20A*, 175–240, 263–297.

Nguyen, T. T., Chen, J. T., Gupta, A. K., & Dinh, K. T. (2003). A proof of the conjecture on positive skewness of generalised inverse Gaussian distributions. *Biometrika*, *90*(1), 245–250.

Nikoloulopoulos, A. K., & Karlis, D. (2008). On modeling count data: a comparison of some well-known discrete distributions. *Journal of Statistical Computation and Simulation*, *78*(3), 437–457.

Ong, S. H. (1988). A discrete Charlier series distribution. *Biometrical Journal*, *30*(8), 1003–1009.

Ong, S. H. (1995). Computation of probabilities of a generalized log-series and related distributions. *Communications in Statistics - Theory and Methods*, *24*(1), 253–271.

Ong, S. H. (1998). A note on the mixed Poisson formulation of the Poisson-inverse Gaussian distribution. *Communications in Statistics - Simulation and Computation*, *27*(1), 67–78.

Ong, S. H., Biswas, A., Peiris, S., & Low, Y. C. (2015). Count distribution for generalized Weibull duration with applications. *Communications in Statistics - Theory and Methods*, *44*(19), 4203–4216.

Ong, S. H., Chakraborty, S., Imoto, T., & Shimizu, K. (2012). Generalizations of non central negative binomial, Charlier series Distributions, and their extensions by Lagrange expansion. *Communications in Statistics - Theory and Methods*, *41*(4), 571–590.

Ong, S. H., & Subarau Muthaloo. (1995). A class of discrete distributions suited to fitting very long-tailed data. *Communications in Statistics - Simulation and Computation*, *24*(4), 929–945.

Ord, J. K., & Whitmore, G. A. (1986). The Poisson-inverse Gaussian distribution as a model for species abundance. *Communications in Statistics - Theory and Methods*, *15*, 853–871.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine*, *50*, 157-172.

Periwal, S. B., Spagna, K., Shahabi, V., Quiroz, J., & Shroff, K. E. (2005). Statistical evaluation for detection of peptide specific interferon-gamma secreting T-cells induced by HIV vaccine determined by ELISPOT assay. *Journal of Immunological Methods*, *305*, 128–134.

Perline, R. (1998). Mixed Poisson distributions tail equivalent to their mixing distributions. *Statistics and Probability Letters*, *38*, 229–233.

Perumean-Chaney, S. E., Morgan, C., McDowall, D., & Aban, I. (2013). Zero-inflated and overdispersed: what's one to do? *Journal of Statistical Computation and Simulation*, *83*(9), 1671–1683.

Phang, Y. N., & Ong, S. H. (2006). Zero-inflated inverse trinomial distribution for modeling count data. In *Proceedings of the 2nd IMT-GT Regional Conference on Mathematics, Statistics and Applications* (pp. 2–7). Penang.

Puig, P., & Valero, J. (2006). Count data distributions: some characterizations with applications. *Journal of the American Statistical Association*, *101*(June 2013), 332–340.

Rao, C. R. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Proceedings of the Cambridge Philosophical Society*, *44*, 50–57.

Rao, C. R. (2005). Score test: Historical Review and Recent Developments. In *Advances in Ranking and Selection, Multiple Comparisons, and Reliability* (pp. 3–20).

Ridout, M., Demetrio, C. G. ., & Hinde, J. (1998). Models for count data with many zeros. *International Biometric Conference*, (December), 1–13.

Ridout, M., Hinde, J., & Demétrio, C. G. (2001). A score test for testing a zero-inflated Poisson regression model against zero-inflated negative binomial alternatives. *Biometrics*, *57*(1), 219–223.

Ridout, M. S., & Besbeas, P. (2004). An empirical model for underdispersed count data. *Statistical Modelling*, *4*, 77–89.

Rigby, R. A., Stasinopoulos, D. M., & Akantziliotou, C. (2008). A framework for modelling overdispersed count data, including the Poisson-shifted generalized inverse Gaussian distribution. *Computational Statistics and Data Analysis*, *53*(2), 381–393.

Rodriguez-Avi, J., Conde-Sanchez, A., Saez-Castillo, A. J., Olmo-Jimenez, M. J., & Martinez-Rodriguez, A. M. (2009). A generalized Waring regression model for count data. *Computational Statistics and Data Analysis*, *53*(10), 3717–3725.

Rose, C. E., Martin, S. W., Wannemuehler, K. a, & Plikaytis, B. D. (2006). On the use of zero-inflated and hurdle models for modeling vaccine adverse event count data. *Journal of Biopharmaceutical Statistics*, *16*(4), 463–481.

Sáez-Castillo, A. J., & Conde-Sanchez, A. (2013). A hyper-Poisson regression model for overdispersed and underdispersed count data. *Computational Statistics and Data Analysis*, *61*, 148–157.

Sankaran, M. (1968). Mixtures by the inverse Gaussian distribution. *Sankhyā: The Indian Journal of Statistics, Series B*, *30*(3-4), 455–458.

Sankaran, M. (1970). The discrete Poisson-Lindley distribution. *Biometrics*, *26*, 145–149.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464.

Seidel, W., Mosler, K., & Alker, M. (2000). A cautionary note on likelihood ratio tests in mixture models. *Annals of the Institute of Statistical Mathematics*, *52*(3), 481–487.

Self, S. G., & Liang, K. Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, *82*(398), 605–610.

Sermutlu, E., & Eyyuboglu, H. T. (2007). A new quadrature routine for improper and oscillatory integrals. *Applied Mathematics and Computation*, *189*(1), 452–461.

Seshadri, V. (1983). The inverse Gaussian distribution: Some properties and characterizations. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, *11*(2), 131–136.

Shaban, S. A. (1981). Computation of the Poisson-inverse Gaussian distribution. *Communications in Statistics - Theory and Methods*, *A10*(14), 1389–1399.

Shaked, M. (1980). On mixtures from exponential families. *Journal of the Royal Statistical Society. Series B (Methodological)*, *42*(2), 192–198.

Shmueli, G., Minka, T. P., Kadane, J. B., & Boatwright, P. (2005). A useful distribution for fitting discrete data: revival of the COM-Poisson. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *54*(1), 127–142.

Sichel, H. S. (1971). On a family of discrete distributions particularly suited to represent long-tailed frequency data. In N. F. Laubscher (Ed.), *Proceedings of the Third Symposium on Mathematical Statistics* (pp. 51–97). Pretoria, South Africa: Council for Scientific and Industrial Research.

Sichel, H. S. (1973). Statistical valuation of diamondiferous deposits. *Journal of the South African Institute of Mining and Metallurgy*, 235–243.

Sichel, H. S. (1974). On a distribution representing sentence-length in written prose. *Journal of the Royal Statistical Society: Series A (General)*, *137*(1), 25–34.

Sichel, H. S. (1975). On a distribution law for word frequencies. *Journal of the American Statistical Association*, *70*(351), 542–547.

Sichel, H. S. (1982a). Asymptotic efficiencies of three methods of estimation for the inverse Gaussian-Poisson distribution. *Biometrika*, *69*(2), 467–472.

Sichel, H. S. (1982b). Repeat-buying and the generalized inverse Gaussian-Poisson distribution. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *31*(3), 193–204.

Skellam, J. G. (1948). A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the sets of trials. *Journal of the Royal Statistical Society. Series B (Methodological)*, *10*(2), 257–261.

Sobol, I. M. (1998). On quasi-Monte Carlo integrations. *Mathematics and Computers in Simulation*, *47*(2-5), 103–112.

Stein, G. Z., & Juritz, J. M. (1988). Linear models with an inverse Gaussian-Poisson error distribution. *Communications in Statistics - Theory and Methods*, *17*(2), 557–571.

Stein, G. Z., Zucchini, W., & Juritz, J. M. (1987). Parameter estimation for the Sichel distribution and its multivariate extension. *Journal of the American Statistical Association*, *82*(399), 938–944.

Stewart, J. A. (1994). The Poisson-lognormal model for bibliometric/scientometric distributions. *Information Processing and Management*, *30*(2), 239–251.

Takeuchi, K. (1976). Distribution of informational statistics and a criterion of model fitting. *Suri-Kagaku (Mathematic Sciences)*, *153*, 12–18. (In Japanese).

Tremblay, L. (1992). Using the Poisson inverse Gaussian in bonus-malus systems. *ASTIN Bulletin*, *22*(I), 97–106.

Trocóniz, I. F., Plan, E. L., Miller, R., & Karlsson, M. O. (2009). Modelling overdispersion and markovian features in count data. *Journal of Pharmacokinetics and Pharmacodynamics*, *36*(5), 461–477.

Ullah, S., Finch, C. F., & Day, L. (2010). Statistical modelling for falls count data. *Accident Analysis and Prevention*, *42*(2), 384–392.

Valkó, P. (2015). Numerical inversion of the Laplace transform. Retrieved August 10, 2016, from www.pe.tamu.edu/valko/public_html/Nil/

Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, *54*(3), 426–482.

Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, *61*(3), 439–447.

Wei, G. C., & Tanner, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association*, *85*(411), 699-704.

Whitmore, G. A. (1979). An inverse Gaussian model for labour turnover. *Journal of the Royal Statistical Society. Series A (General)1*, *142*(4), 468–478.

Williams, M. S., & Ebel, E. D. (2012). Methods for fitting the Poisson-lognormal distribution to microbial testing data. *Food Control*, *27*(1), 73–80.

Willmot, G.E (1986). Mixed compound Poisson distributions. *ASTIN Bulletin*, *16*(3), 59–79.

Willmot, G.E. (1990). Asymptotic tail behaviour of Poisson mixtures with applications. *Advances in Applied Probability*, *22*(1), 147–159.

Willmot, G.E., & Sundt, B. (1989). On posterior probabilities and moments in mixed Poisson processes. *Scandinavian Actuarial Journal*, *14*, 139-146.

Winkelmann, R. (1995). Duration dependence and dispersion in count-data models. *Journal of Business & Economic Statistics*, *13*(4), 467-474.

Winkelmann, R., & Zimmermann, K. F. (1995). Recent developments in count data modelling: Theory and application. *Journal of Economic Surveys*, *9*(1), 1-24.

Wu, C. F. J. (1983). On the convergence properties of the EM algorithm. *The Annals of Statistics*, *11*(1), 95–103.

Xekalaki, E. (1983). The univariate generalized Waring distribution in relation to accident theory: proneness, spells or contagion? *Biometrics*, *39*(4), 887–895.

Xekalaki, E. (2014). On the distribution theory of over-dispersion. *Journal of Statistical Distributions and Applications*, *1*(19), 1–22.

Yau, K. K. W., Wang, K., & Lee, A. H. (2003). Zero-inflated negative binomial mixed regression modeling of over-dispersed count data with extra zeros. *Biometrical Journal*, *45*(4), 437–452.

Yip, K. C. H., & Yau, K. K. W. (2005). On modeling claim frequency data in general insurance with extra zeros. *Insurance: Mathematics and Economics*, *36*(2), 153–163.

Zakerzadeh, H., & Dolati, A. (2009). Generalized Lindley distribution. *Journal of Mathematical Extension*, *3*(2), 13–25.

Zeviani, W. M., Ribeiro Jr, P. J., Bonat, W. H., Shimakura, S. E., & Muniz, J. A. (2014). The gamma-count distribution in the analysis of experimental underdispersed data. *Journal of Applied Statistics*, (October), 1–11.

Zhu, R., & Joe, H. (2009). Modelling heavy-tailed count data using a generalised Poisson-inverse Gaussian family. *Statistics and Probability Letters*, *79*(15), 1695–1703.

LIST OF PUBLICATIONS AND PAPERS PRESENTED

Ong, S. H., Biswas, A., Peiris, S., & Low, Y. C. (2015). Count distribution for generalized Weibull duration with applications. *Communications in Statistics - Theory and Methods*, *44*(19), 4203–4216.

Low, Y.C., & Ong, S.H. (2016). Count distribution for mixture of two exponentials as renewal process duration with applications. Paper presented at the 2nd International Conference for Mathematical Sciences and Statistics, 26-28 January 2016, Kuala Lumpur, Malaysia. Full paper to be published in the conference proceedings.

Low, Y.C., Ong. S.H., & Gupta, R.C. (2014). Modelling overdispersion and heavy tails by the generalized Sichel distribution. Paper presented at the 22nd National Symposium on Mathematical Sciences, 24-26 November 2014, Shah Alam, Malaysia.